

## University of Groningen

### Symposiumbundel het web gevangen

den Hollander, Frank; Voerman, Gerrit

DOI:  
[10.13140/2.1.1810.4643](https://doi.org/10.13140/2.1.1810.4643)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2003

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*  
den Hollander, F., & Voerman, G. (editors) (2003). *Symposiumbundel het web gevangen: het archiveren van de websites van de Nederlandse politieke partijen*. Documentatiecentrum Nederlandse Politieke Partijen. <https://doi.org/10.13140/2.1.1810.4643>

#### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

#### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

SYMPOSIUMBUNDEL

# het web gevangen

HET  
ARCHIVEREN  
VAN DE  
WEBSITES  
VAN DE  
NEDERLANDSE  
POLITIEKE  
PARTIJEN

**archi  
POL**

Het web gevangen



symposiumbundel

# het web gevangen

het  
archiveren  
van de  
websites  
van de nederlandse  
politieke partijen

*redactie*  
*Frank den Hollander en Gerrit Voerman*

Uitgave: Universiteitsbibliotheek Groningen  
Documentatiecentrum Nederlandse Politieke Partijen

[www.archipol.nl](http://www.archipol.nl)  
[info@archipol.nl](mailto:info@archipol.nl)

Vormgeving: René Lapoutre

Druk: Raster Prepress, Groningen

ISBN 90-367-1760-4

Toen de eerste websites verschenen, werden deze gezien als een interessant nieuw verschijnsel, een soort tussenvorm tussen de wereld van het geschreven woord en de visuele wereld van de televisie. Websites hebben met de eerstgenoemde wereld gemeen dat zij veel tekst bevatten en met de laatste wereld dat zij virtueel zijn en daarom vluchtig. Zij vormen geen statisch medium, maar een platform waarop wijzigingen met de regelmaat van de dag, week of maand worden aangebracht. Wie met zijn tijd mee wil gaan, zorgt er voor dat zijn website na verloop van tijd geheel wordt herzien, gemoderniseerd en voorzien van de nieuwste snufjes op het gebied van de informatietechnologie.

Pas geleidelijk is het besef doorgedrongen dat het de moeite waard is om de ontwikkeling van de vormgeving en vooral de inhoud van websites te bestuderen, dat zij van waarde kunnen zijn of worden als historisch bronnenmateriaal, en dat het de moeite waard is bepaalde websites voor de toekomst te bewaren. Het is geen wonder dat dit besef vooral ontstond in documentatiecentra, bibliotheken en archieven. Zij dragen immers een bijzondere verantwoordelijkheid voor het vastleggen van de geschiedenis.

Bij de Rijksuniversiteit Groningen (RUG) bestaat al sinds jaar en dag de gelukkige situatie van intensieve contacten tussen het Documentatiecentrum Nederlandse Politieke Partijen (DNPP) en de Universiteitsbibliotheek (UB). Beide instellingen zien het als een van hun opdrachten om faciliteiten te ontwikkelen voor het bewaren, of beter gezegd het archiveren van websites, in het bijzonder die van de Nederlandse politieke partijen.

Deze vruchtbare samenwerking leidde tot het projectvoorstel 'Archipol. Het archiveren van websites van Nederlandse politieke partijen'. De ontwerpers van dit voorstel, dr. G. Voerman, directeur van het DNPP en A.K. Keyzer, stafmedewerker elektronische dienstverlening van de UB, dienden dit project met succes in bij de stuurgroep Innovatie Wetenschappelijke Informatievoorziening (IWI) van de Stichting SURF en het College van Bestuur van de RUG. Dankzij subsidies van hun kant, aangevuld met eigen middelen van het DNPP en de UB, kon het project tot een goed einde worden gebracht. Daarbij speelden andere collega's een belangrijke rol. Zo programmeerde H.C.G. Druiven, ICT-deskundige en webontwikkelaar bij de UB, met veel inventiviteit de benodigde software. F. J. den Hollander, webmaster en voorlichtingsfunctionaris, en M. S. van Delden, medewerker elektroni-

sche dienstverlening van de UB, verleenden belangrijke steun, evenals W. Koster en K. Visser van het Rekencentrum van de RUG. Hun allen zeg ik hierbij van harte dank.

Terugziend kunnen we met voldoening constateren dat het project tot een goed einde is gebracht. Methoden en technieken zijn voorhanden om het web te vangen, niet alleen de websites van de Nederlandse politieke partijen, maar in principe alle sites. Zo werd ook de omvangrijke website van de RUG gearchiveerd.

Het DNPP en de UB zijn van mening dat zij een bijzonder interessant project met succes hebben voltooid. Zij willen U door middel van deze publicatie graag deelgenoot maken van de bereikte resultaten.

6

Dr. Alex C. Klugkist

*Bibliothecaris van de Rijksuniversiteit Groningen*



<b>Woord vooraf</b>	<b>5</b>
<i>Alex Klugkist</i>	
<b>1. Inleiding</b>	<b>9</b>
<i>Gerrit Voerman</i>	
<b>2. Het archiveren van websites</b>	<b>13</b>
<i>Frank den Hollander en Gerrit Voerman</i>	
<b>3. International initiatives in archiving the Web</b>	<b>19</b>
<i>Julien Masanès</i>	
<b>4. Het Archipol-project</b>	<b>25</b>
<i>Henk Druiven, Frank den Hollander, André Keyzer en Gerrit Voerman</i>	
<b>5. Juridische aspecten van het archiveren van websites</b>	<b>35</b>
<i>Wilma Mossink</i>	
<b>6. Migratie</b>	<b>45</b>
<i>André Keyzer</i>	
<b>Slot</b>	<b>49</b>
<i>Gerrit Voerman</i>	
<b>Publicaties over het Archipol-project</b>	<b>51</b>
<b>Het symposium: een foto-impressie</b>	<b>53</b>
<b>Personalia</b>	<b>55</b>



# 1 | INLEIDING

*Gerrit Voerman*

9

Op 25 april 2002 opende J. Wallage, burgemeester van Groningen en kort daarvoor nog voorzitter van de Commissie Toekomst Overheidsdocumentatie, het campagne-archief van het Archipol-project. Archipol staat voor ARCHief van websites van POLitieke partijen in Nederland. In het kader van dit project worden sinds april 2001 de websites van de Nederlandse politieke partijen gearchiveerd. Ten tijde van de campagne voor de Tweede-Kamerverkiezingen in het voorjaar van 2002 werden niet alleen de sites van alle deelnemende politieke partijen, maar ook die van hun kandidaten en andere verkiezingssites in het zogeheten 'campagne-archief' opgenomen. In dit project werken het Documentatiecentrum Nederlandse Politieke Partijen (DNPP) en de Universiteitsbibliotheek van de Rijksuniversiteit Groningen samen.

De voorgeschiedenis van het Archipol-project gaat terug tot de Tweede-Kamerverkiezingen van mei 1998. In die tijd werd op het DNPP onderzoek gedaan naar de manier waarop de politieke partijen in de verkiezingscampagne gebruik maakten van internet. Vrijwel alle partijen hadden in die tijd een website, met informatie over onder meer het verkiezingsprogramma, de kandidaten en de campagne. Een groot probleem bij het onderzoek vormde het vastleggen van de snel veranderende informatie op de sites. Wanneer pagina's van de site werden uitgeprint, dan kwam zelden alles op papier te staan wat op dat moment op het scherm van de monitor te zien was. Van de ene partijsite werd alleen het frame uitgeprint, en van de andere alleen de tekst. Bij een derde vielen de foto's weg. De optie 'print screen' leverde in het meest gunstige geval slechts een deel van de gewenste pagina op.

Na overleg met de automatiseringsafdeling van de Universiteitsbibliotheek bleek er eigenlijk maar één praktische en werkzame oplossing te bestaan om ten behoeve van het onderzoek over de websites te beschikken, namelijk ze allemaal in hun geheel te downloaden. Een gezamenlijk project was hiermee geboren.

De websites van de Nederlandse politieke partijen zouden moeten worden bewaard voor de toekomst, niet alleen als bronnenmateriaal voor latere studie, maar ook omdat zij steeds vaker informatie bevatten die niet meer in gedrukte vorm verschijnt.

Het archiveringsproject startte in 2000, nadat subsidie was verkregen van het College van Bestuur van de Rijksuniversiteit Groningen en van de stuurgroep Innovatie Wetenschappelijke Informatievoorziening (IWI) van de Stichting SURF. Daarnaast bood de Universiteitsbibliotheek royale ondersteuning. Aan het project werd meegewerkt door André Keyzer, Frank den Hollander, Henk Druiven en Mischa van Delden, en vanuit het Rekencentrum door Willem Koster en Kees Visser. Ik wil hen allen hartelijk danken voor de bijdrage die zij hebben geleverd aan het welslagen van dit project.

In deze bundel zijn onder andere de bijdragen opgenomen van het symposium 'Het web gevangen', dat de Universiteitsbibliotheek en het DNPP organiseerden ter gelegenheid van de opening van het campagne-archief op 25 april 2002. Deze bijeenkomst was geheel gewijd aan het archiveren van websites.

De verzamelingsstrategie, de bewerkingen die de downloads ondergaan en de ontsluiting van de gearchiveerde sites in het Archipol-archief, kwamen uiteraard uitgebreid aan bod tijdens dit symposium. In deze bundel is aan deze thema's een hoofdstuk gewijd.

Julien Masanès (Bibliothèque Nationale de France) gaat in op het onderwerp van het archiveren van websites en de diverse internationale initiatieven op dit terrein. Hij bespreekt twee modellen voor archiveren: de selectieve benadering en de *snapshot*. Bij de selectieve benadering wordt een van tevoren uitgekozen aantal sites periodiek gearchiveerd; dit is de methode die ook in het Archipol-project wordt toegepast. Bij een snapshot wordt een 'momentopname' gemaakt van het hele internet (voor zover dat althans mogelijk is), of bijvoorbeeld van het hele Zweedse domein binnen dat internet.

Wilma Mossink (Open Universiteit) behandelt de auteursrechtelijke aspecten van het archiveren van websites, vooral in het licht van de nieuwe Auteurswet die nu voor advies bij de Raad van State ligt.

Inmiddels maken steeds meer mensen gebruik van het Archipol-archief. Vooral na de dramatische ontwikkelingen in de aanloop naar de Tweede Kamerverkiezingen van 15 mei bleek hoe belangrijk het was om te kunnen onderzoeken hoe de verschillende partijen gebruik maakten van hun websites. Zo waren na de moord op Pim Fortuyn enkele sites volledig uit de lucht gehaald, op één sobere voorpagina na. Op de vele sites en nevensites van de Lijst Pim Fortuyn werden uitgebreide condoleanceregisters opgezet en vonden emotionele discussies plaats. Deze websites zijn, evenals vele fansites en andere webuitingen over deze gebeurtenissen, gedownload en gearchiveerd door Archipol.

Bij het ter perse gaan van deze bundel heeft zich alweer een nieuwe uitdaging aangediend: nu er voor de tweede keer binnen een jaar Tweede-Kamerverkiezingen worden gehouden, is het Archipol-archief onmiddellijk uitgebreid met de meest actuele verkiezingssites en de websites van de voor deze verkiezingen nieuw opgerichte partijen.

11



De voorpagina van de Lijst Pim Fortuyn (LPF) op 13 mei 2002



## 2 | HET ARCHIVEREN VAN WEBSITES

*Frank den Hollander en Gerrit Voerman*

### **:: Verdwijnend digitaal erfgoed**

“Pantha rhei”, was één van de uitgangspunten van de filosoof Heraclitus. Deze observatie van het leven in Griekenland in de zesde eeuw voor Christus geldt zo’n 2500 jaar later zeker ook voor het World Wide Web. Ook op het web is alles aan voortdurende verandering onderhevig. Het web breidt zich gigantisch snel uit: elke maand komen er miljoenen nieuwe pagina’s bij. Het aantal sites zou jaarlijks met miljoenen toenemen en in juli 2002 162 miljoen bedragen.<sup>1)</sup> Deze fabelachtige groei voltrekt zich ondanks het feit dat ook veel sites weer verdwijnen: de gemiddelde levensduur van een site wordt geschat op 75 dagen. Zij worden vaak maar op één lokatie op het web aangeboden; wanneer de aanbieder er om wat voor reden dan ook mee stopt, is de site voor altijd verloren. In dit dynamische proces van opkomst en ondergang vormen de meeste bestaande sites geen statisch rustpunt. Zij veranderen immers voortdurend: een paar seconden na een bezoek aan een site kan deze alweer gewijzigd zijn, doordat de aanbieder informatie heeft toegevoegd, of een bezoeker een bericht heeft achtergelaten.

Vanwege dit continue proces van verandering van het web, zou men mogen verwachten dat websites en hun vroegere verschijningsvormen van meet af aan bewaard zouden worden. Het duurde echter geruime tijd sinds het web aan het begin van de jaren negentig zijn intrede had gedaan, voordat het zover was.<sup>2)</sup> Hierdoor is inmiddels een substantieel deel van dit digitale erfgoed verloren gegaan. Voor toekomstig onderzoek naar de ontwikkeling van de ‘virtuele’ wereld van het web en zijn relatie met de ‘werkelijke’ samenleving, ontbreekt het dan ook aan bronnenmateriaal. “Internet dreigt een gat in de geschiedschrijving te veroorzaken”, zo constateerden Amerikaanse historici op een congres in 1998 niet ten onrechte.<sup>3)</sup>

### **:: Internationale archiveringsinitiatieven**

Hoewel er dus in het algemeen weinig oog was voor het belang van de bewaring van websites, werd een jaar of vijf geleden hier en daar een voorzichtig begin gemaakt met het op meer systematische wijze archiveren van sites (zie voor de internationale archiveringsinitiatieven ook de hierna volgende bijdrage van Julien Masanès). Eén van de pioniers was de *National Library of Australia*, dat in juni 1996

het Pandora-project (Preserving and Accessing Networked Documentary Resources in Australia) opzette. Belangrijk geachte Australische on-line publicaties worden in het kader van Pandora gearchiveerd. Onder meer worden enkele websites van Australische politieke partijen bewaard.<sup>4)</sup> Een paar maanden later, in september 1996, begon de Koninklijke Bibliotheek van Zweden met het zogeheten 'The Kulturarw3 Project'. Doel is zoveel mogelijk te archiveren van het Zweedse deel van internet, dat wil zeggen alle URL's met de extensie *se*. Hiervan zijn sindsdien enkele *snap-shots* gemaakt, momentopnames waarbij in totaal ongeveer 126.000 websites zijn gearchiveerd. Als gevolg van juridische obstakels is het digitale archief niet vrij voor het publiek toegankelijk.<sup>5)</sup> Het Zweedse project heeft inmiddels bij de Nationale Bibliotheek in Finland navolging gevonden. In de zomer van 2002 is de eerste download van het Finse deel van het internet voltooid.

In Nederland richten twee instanties zich op de archivering van delen van het internet. De Koninklijke Bibliotheek is begonnen met de inrichting van het Depot van Nederlandse Elektronische Publicaties (DNEP).<sup>6)</sup> Hierin zullen naast *off line* digitale publicaties als cd-rom's in de toekomst ook *on line* publicaties als elektronische tijdschriften, boeken en artikelen worden opgenomen. Het Nederlands Uitgeversverbond heeft inmiddels ingestemd met een regeling waarbij deze publicaties bij het DNEP worden gedeponeerd. Verder kunnen ook bepaalde webdocumenten worden ondergebracht in het Depot; complete sites worden niet verzameld, noch archieven van discussielijsten.<sup>7)</sup>

Voor dat laatste heeft het Internationaal Instituut voor Sociale Geschiedenis (IISG) grote belangstelling. Actiegroepen en sociale bewegingen ontdekten al vroeg het internet als een goedkope en effectieve manier om hun standpunten te verspreiden en steun te mobiliseren. Zoals het IISG vroeger de pamfletten en brochures van deze groeperingen verzamelde, collectioneert het Instituut nu hun berichten in nieuwsgroepen. In 1995 werd onder de naam 'Occasio' begonnen met het archiveren van bijna duizend internet-nieuwsgroepen van de Association for Progressive Communications. Het archief, dat onder meer veel materiaal bevat over de burgeroorlogen in het voormalige Joegoslavië en de opstand tegen de toenmalige president Soeharto in Indonesië, telde in 2002 ongeveer 2,3 miljoen berichten.<sup>8)</sup>

Het meest grootschalige project werd ontwikkeld door de Amerikaanse computerprogrammeur Brewster Kahle. Hij is de oprichter van het in San Francisco gevestigde 'Internet Archive', dat officieel sinds de zomer van 1996 – zoals de naam al aangeeft – het gehele internet tracht te archiveren, van nieuwsgroepen tot home pages. Hiervoor wordt gebruik gemaakt van *web crawling robots*: programma's die via externe links van sites andere sites opzoeken en ze allemaal in hun geheel downloaden. Op deze wijze wordt een momentopname van het internet gemaakt. In



oktober 2001 had het Archive circa tien miljard ongespecificeerde webpagina's verzameld – overwegend html-bestanden overigens.<sup>9)</sup> In de herfst van 2001 kwam deze collectie via de zogeheten 'Way Back Machine' online beschikbaar.

## **:: Het belang van de bewaring van de sites van de Nederlandse politieke partijen**

Omdat ook de bewaring van de sites van de politieke partijen in Nederland zeer te wensen overliet, begon het Documentatiecentrum Nederlandse Politieke Partijen (DNPP) – zoals reeds vermeld – in 2000 samen met de Universiteitsbibliotheek van de Rijksuniversiteit Groningen met het Archipol-project (ARCHIef van websites van POLitieke partijen in Nederland). In tegenstelling tot het Amerikaanse Internet Archive richt het Achipol-project zich op een beperkt aantal websites. Sinds de Nederlandse politieke partijen hun debuut maakten op het web, hebben de meeste hun site drie à vier keer compleet aangepast. Van de oudere versies is vrijwel niets meer over (behalve een paar die deel uitmaakten van het Internet Archive, zoals bleek na de komst van de Way Back Machine). Het gevolg hiervan is dat de eerste stappen van de Nederlandse politieke partijen op internet niet meer kunnen worden getraceerd.

Politieke partijen waren redelijk vroeg op het web te vinden. In januari 1994 was GroenLinks de eerste die met een website begon.<sup>10)</sup> In een beweging van links naar rechts volgden de andere partijen: de PvdA in november 1994, D66 en het CDA medio 1995, en de VVD in het voorjaar van 1997. Hekkensluiter was de SGP in de herfst van 2000.<sup>11)</sup>

Aan de vooravond van de Tweede-Kamerverkiezingen van mei 1998 hadden alle partijen – behalve de SGP – een site. Veel aanloop hadden zij niet tijdens de verkiezingscampagne. Naar schatting bezochten maximaal 100.000 personen de maand voor de stembusdag een partij-site.<sup>12)</sup> Vier jaar later echter, in mei 2002, lag dit aantal spectaculair hoger. Ongetwijfeld zal dat mede te maken hebben gehad met de bijzondere omstandigheden waaronder de verkiezingscampagne plaats had, zoals de electorale opkomst van de rechts-populistische Pim Fortuyn en zijn gewelddadige dood op 6 mei. In de laatste tien dagen van de verkiezingscampagne werden de sites van de in de Tweede Kamer vertegenwoordigde partijen en die van de Lijst Pim Fortuyn (LPF) meer dan een miljoen keer bezocht. Afgezien daarvan bracht de Stemwijzer meer dan twee miljoen stemadviezen uit (tegen 6.500 in 1998).<sup>13)</sup>

Deze cijfers maken duidelijk dat in kwantitatief opzicht de sites voor de partijen als een schakel in de communicatie met de achterban belangrijker zijn geworden dan ooit te voren. De sites spelen al enige tijd een grote rol in de informatie-

voorziening aan kiezers en leden – en dat niet alleen binnen, maar ook buiten het electorale seizoen. Verkiezingsprogramma's, persberichten, achtergronden van kandidaten, artikelen uit periodieken – het verschijnt niet alleen in druk, maar het is doorgaans ook allemaal op de site te vinden.

Bovendien hebben partijen steeds meer de neiging om bepaalde informatie uitsluitend digitaal aan te bieden, omdat dat een stuk sneller en veel goedkoper gaat. De veranderende inhoud van de ledenbladen getuigt hiervan. Deze periodieken ontwikkelen zich meer tot magazines, waarin niet langer de wat saaie partijorganisatorische informatie wordt vermeld (zoals spreekbeurten, agenda's van partijbijeenkomsten en dergelijke). D66 bijvoorbeeld drukt de stukken voor het partijcongres niet langer in het ledenblad af, maar plaatst die uitsluitend op haar website.

16

Overigens evolueert de partij-site niet alleen door deze verschuiving van gedrukte naar digitaal aangeboden informatie tot exclusieve publicatieplaats voor bepaalde categorieën informatie. Deze bijzondere positie krijgt een website ook doordat partijen – met name in verkiezingscampagnes – met speciale internetredacties werken, die uitsluitend berichten voor de site schrijven.

Naast het op snelle en relatief goedkope wijze aanbieden van informatie die bovendien makkelijk up to date te houden is, zijn sites ook om andere redenen voor partijen in toenemende mate interessant. Hoewel de meeste sites nog altijd vooral bedoeld zijn om partij-informatie te verspreiden – dus nogal *top down* zijn opgezet – mag verwacht worden dat de interactieve mogelijkheden de komende tijd meer gewicht zullen krijgen. In de Verenigde Staten bijvoorbeeld worden websites gebruikt om campaignemedewerkers te werven en financiële steun binnen te krijgen. Daarnaast zullen sites een grotere rol gaan spelen in de meningsvorming binnen een partij. Gezien hun voortdurend ledenverlies hebben de meeste partijen tegenwoordig de grootste moeite om leden aan zich te binden en hen te activeren. De nieuwe vormen van digitale participatie die de laatste tijd beschikbaar zijn gekomen, bieden de partijen eigentijdse mogelijkheden om met name jongeren meer bij de politiek te betrekken – zonder dit nu overigens meteen als het panacee te willen zien.<sup>14)</sup>

In Nederland experimenteren partijen daar momenteel ook mee. Het CDA neemt hierbij het voortouw. De christen-democraten gebruikten de partij-site om ideeën te genereren voor het program voor de parlementsverkiezingen van 2002. Duizenden bezoekers van de site deden hier aan mee. Het aantal bruikbare ideeën was niet bijzonder groot, maar de partij kreeg wel een indruk wat er onder haar aanhang leefde. Daarnaast werd een intern 'CDA-net' ingericht waartoe alleen partijleden (door middel van hun lidmaatschapsnummer) toegang hebben. Het was de

bedoeling dat deze digitale afdeling in de 'normale' partijorganisatie werd geïntegreerd, maar daarvan is voorlopig afgezien vanwege problemen bij de praktische uitvoering.

Sites spelen dus een steeds grotere rol binnen partijen – hetzij in de informatievoorziening dan wel bij het bieden van participatiemogelijkheden. Naast de sites die direct onder de verantwoordelijkheid van de partijleiding vallen, hebben ook bepaalde onderdelen van partijen (bijvoorbeeld partijcommissies), nevenorganisaties (jongeren, vrouwen) en Tweede-Kamerkandidaten websites geopend. In een enkel geval zijn er ook sites gekomen voor bijzondere categorieën leden, zoals de kadersite van GroenLinks, primair bedoeld voor afdelingsbesturen, raads- en statenleden, wethouders en burgemeesters van deze partij. Naast deze min of meer 'officiële' sites hebben de afgelopen jaren ook 'officieuze' sites van (ad hoc) presiegroepen binnen de partijen hun intrede gedaan. Binnen D66 bijvoorbeeld hebben groepen ontevreden leden diverse keren met behulp van een website geprobeerd de partijlijn te beïnvloeden – niet zonder succes. Ook in de in het voorjaar van 2002 door Fortuyn opgerichte LPF is naar dit middel gegrepen door opposanten van het partijbestuur. Bij interne verkiezingen wordt eveneens gebruik gemaakt van sites. Bij verkiezing van een nieuwe voorzitter van de PvdA in het voorjaar van 2001 hadden alle vier de kandidaten een eigen site. In de interne besluitvorming binnen partijen beginnen sites dus ook een rol te spelen.

De hierboven aangeduide ontwikkelingen vormden voor het DNPP een belangrijk motief om de sites te gaan archiveren. In de eerste plaats zou het niet vastleggen van de digitale uitingen van de partijen leiden tot een – steeds groter wordend – hiaat in de verzameling van partijpublicaties van het Documentatiecentrum. De kans dat bepaalde informatie die alleen nog maar op de website verschijnt verloren gaat, is immers groot. Naast deze bibliotheek-technische redenen dienen websites ook bewaard te worden vanwege hun toenemende belang als communicatie-, participatie-, mobilisatie- of pressiemiddel. De bestudering van politieke partijen en de wijze waarop zij intern en extern communiceren, zal het in de toekomst moeilijk kunnen stellen zonder gearchiveerde sites.

1 Internet Software Consortium; [www.isc.org/ds/WWW-200207/index.html](http://www.isc.org/ds/WWW-200207/index.html).

2 Zie C. Caey, 'The Cyberarchive: A Look at the Storage and Preservation of Web Sites', in: College & Research Libraries, juli 1998, 304-310.

3 De Volkskrant, 30 januari 1999.

4 Zie [pandora.nla.gov.au/index.html](http://pandora.nla.gov.au/index.html).

- 5 Zie A. Arvidson en F. Lettenström, 'The Kulturarw3 Project - the Swedish Royal Web Archive', in: *The Electronic Library*, 16 (1998), 2 (April), 105-108; zie ook [www.Kulturarw3.kb.se/html/projectdescription.html](http://www.Kulturarw3.kb.se/html/projectdescription.html).
- 6 Zie [www.kb.nl/kb/resources/frameset\\_kb.html?/kb/dnp/dnep.html](http://www.kb.nl/kb/resources/frameset_kb.html?/kb/dnp/dnep.html)
- 7 'Symposium "Het eeuwige leven"', in: *Historia & Informatica*, 5 (1998), 1 (apr.), 6-7; G.C. Noordermeer, 'Depot van Nederlandse elektronische publicaties', in: *HG-Nieuws*, 4 (1997), 1 (mrt.), 4-5; idem, 'Depot van Nederlandse Elektronische Publicaties', in: *Informatie Professional*, 1998, nr. 2, 22-24.
- 8 J. Quast, 'OCCASIO Digital Social History Archive', in: *Historia & Informatica*, 5 (1998), nr. 2; zie ook idem, 'IISG: Occasio Archief nu toegankelijk via het Web', in: idem, 6 (1999), 3 (dec.), 1; en idem, 'Het internetarchief van het IISG', in: *Archievenblad*, 104 (2000), 7 (sep.), 16-17. Zie voor meer informatie: [www.iisg.nl/occasio](http://www.iisg.nl/occasio).
- 9 Zie onder meer B. Kahle, 'Preserving the Internet', in: *Scientific American*, 1997, nr. 3; M. Cunningham, 'Brewster's millions', in: *The Irish Times* (webversie), 27 januari 1997; interview met Kahle in: *RLG DigiNews*, 6 (2002), 3 (15 juni), [www.rlg.org/preserv/dig-inews/diginews6-3.html](http://www.rlg.org/preserv/dig-inews/diginews6-3.html). Zie voor de website van het Internet Archive [www.archive.org](http://www.archive.org).
- 10 S. Ward en G. Voerman, 'New media and new politics. Green Parties, Intra-party democracy and the Potential of the Internet (an Anglo-Dutch Comparison)', in: *Jaarboek 1999 Documentatiecentrum Nederlandse Politieke Partijen*, Groningen, 2000, 192-215; zie [www.ub.rug.nl/eldoc/dnpp/1999/h9.pdf](http://www.ub.rug.nl/eldoc/dnpp/1999/h9.pdf).
- 11 G. Voerman en J.D. de Graaf, 'De web-sites van de Nederlandse politieke partijen, 1994-1998', in: *Jaarboek 1997 DNPP*, Groningen, 1998, 238-269; zie [www.ub.rug.nl/eldoc/dnpp/1997/011.pdf](http://www.ub.rug.nl/eldoc/dnpp/1997/011.pdf).
- 12 G. Voerman, 'Elektronisch folderen: de digitale campagne', in: Ph. van Praag en K. Brants, *Tussen beeld en inhoud. Politiek en media in de verkiezingen van 1998*, Amsterdam, 2000, 206.
- 13 Zie *StemWijzer 2002. Reacties van publiek en politiek*, z.pl. [Amsterdam], z.j. [2002].
- 14 De mogelijkheden om via websites groepen te bereiken die geen al te grote belangstelling voor de politiek hebben lijken, overigens tamelijk beperkt; zie M. Boogers, G. Voerman en E. Andersson, *Enquête Politiek & Internet. Resultaten van een online-enquête onder bezoekers van politieke websites tijdens de verkiezingscampagne van 2002*, Tilburg/Groningen, 2002 (te vinden op [www.ub.rug.nl/dnpp/online-enquete.pdf](http://www.ub.rug.nl/dnpp/online-enquete.pdf)).



# 3 | INTERNATIONAL INITIATIVES IN ARCHIVING THE WEB

*Julien Masanès*

## **:: Introduction**

In this lecture I will discuss methods and approaches for webarchiving that have been applied in different countries for more than six years now, the problems that arise in the process, and what we can do to try to overcome these problems.

First I will say a few words about archiving in the Internet environment. Subsequently, we will deal with the two main approaches that have been taken by different countries. Then we will see how these two approaches can mix with one another, and what kind of new tools are needed for that.

19

## **:: Preserving web content**

I will focus on global archiving of the Internet environment, not the event-related or thematic collection, for instance of political sites. The global national initiatives to archive at least a national domain for Sweden, for France, and so on, are in fact prolongations of what has been done with the traditional published material for a long time, whether it is called a legal deposit or a copyright deposit. In France, for instance, legal deposit is intended to preserve a large range of material so researchers have the possibility to find a very specific book one hundred or two hundred years later. This long-term and global archiving is the main subject of my lecture.

On the Internet, we have to deal with unfiltered content: everybody can 'publish' something on the net. This makes a great difference for us, because we have always been working with what has been selected and transformed by publishers. On the Internet we have publishers too, of course, but we also have a large amount of material which has not gone through a publishing process and that is quite new, because it means we have to assume a selection. This selection must not be done in the same way as we buy books or borrow at a university library, because the idea is not to archive the best thing or the most well known at that moment. The idea is to archive a larger amount of things to be able to find things or material that would not have been selected at the time of archiving. When we are in the reading room at the National Library people do not come in and ask: 'I want a book by Flaubert'. You do not need to read Flaubert at the National Library, except if you want to check the original edition. You would want to read for instance a very, very

specific newspaper from the nineteenth century on this specific town that would never have been selected by a librarian at that time.

Another main difference with traditional material is that the content on the Internet is dynamic. People can change their website during the night and you are not aware of it. And there is another difference, which this time is a positive one: we can, to a certain extent of course, gather content online directly. For instance, we can copy sites from the National Library directly.

To face this changing environment, two approaches have been adopted. The first one is the selective approach. Our colleagues from Canada and Australia<sup>1)</sup> have initiated this approach now six years ago. They select a few hundred sites, copy and catalogue them.

20

The second one is the snapshot approach. For instance, the Royal Library of Sweden<sup>2)</sup> is doing this, and also the Internet Archive.<sup>3)</sup> What we mean by a snapshot is a global capture of all sites the robot can gather in one moment. For a few weeks or months the robot goes from site to site, following links, and archiving everything it encounters. That makes a kind of image or snapshot of the whole web at that moment.

### **:: The selective approach**

When archiving a site one can use the push model or the pull model. The pull model is: you go to a site and with a little robot copy all files of this particular site without going out of it (contrary to the snapshot). You make a so-called mirror copy of the site. In the push model you ask the site's webmaster to send you a carrier with all the files on it.

Whether you use the one method or the other, you can only just archive a few hundred sites, or maybe a few thousand, because you have to select them, to catalogue them, to collect them, and so on. It acquires a large amount of work done manually. Another problem is that you have to define a selection policy, which is quite difficult, because it is hard to define which site will be of interest for researchers in two hundred years.

You also have to define for each site the archiving frequency, which of course depends on the analysis of the site you made. You may consider that this site is changing very often or not. And if it does change, it can very well keep all its older content in it, so that if you come again in one month or in six months you will still have the whole content there.

The strength of this method is first that it makes it possible to follow the

development of each site in time. That means you are able to adapt the frequencies, the kind of copy and so on for each site.

Another strength is that you can check each site individually and if it cannot be copied online you can try to contact the webmaster and ask him to provide you directly with the files. That is a good point for this approach.

The main weakness of this approach is that it cannot be applied at a large scale. For instance, if we could archive three thousand sites in France in this way, that would still be just only one percent of the general web. And if you just archive one percent of the content, that is a fundamental difference with traditional legal deposit, where you have almost everything. An other point is that you cannot navigate from site to site, because you just have a few sites archived.

## **:: The snapshot approach**

21

The snapshot method is based on automatic tools: robots or spiders, as we call them. These are the same kind of tools that are used by the search engine. First you feed them with a few hundreds URLs, preferably URLs that have many outgoing links. The robot fetches these pages, parses the html-code, extracts each hypertext link and puts them in a queue. Then it will do the same with each link that has been queued. Thus, it will go from link to link through the whole web. That is the way search engines explore the web. The most powerful search engine at the moment, Google, has a database with 2.5 billion URLs.

The difference between archiving robots and search engine robots is that search engines generally get non-html documents and they just index pages but don not archive them. Archiving robots fetch every page, every linked image and any kind of file they can get, just as it is.

Robots can get everything in what we call the surface-web or visible web. This expression is not a very convenient one, but it means exactly: what is visible for the robot. You as a human might be surfing on the deep web without even noticing it. But the robot will not be surfing through this space, because it cannot. For instance, when you have an html form, you have to enter a title or something. Of course a robot cannot do this, so it will archive just the form and nothing more. Everything that is behind is the deep web or invisible web and will not be in your archive if you just use robots.

Access to the archive can be done by the URL of the site you want to see or by a full text index like in the Swedish archive or Internet Archive, just as in a search engine. You can navigate through the archive just as you navigate through the web. It has the same form, because links point from page to page in the archive.

What are the strengths of this snapshot approach? Of course, the main advantage is the coverage of the domain. It gives a much more representative result than what is done when you select the sites one by one. And the archive is fully navigable.

One important weakness of this approach, however, is the insufficient dynamic follow-up of sites. There are just two snapshots, which means only two versions of a site per year. And part of the web, and a quite interesting part at that, is out of reach for the robot – the deep web. That means the deep web is not archived by these robots. In fact there is some kind of selection but only on a technical basis, because what is technically the deep web will not be archived.

## **:: An integrative approach**

22

From the discussion between national libraries<sup>4)</sup>, some consensus emerges that the best thing to do is some kind of mixture of the two approaches. At the National Library of France we have been working in that direction for more than two years now, in a two-tracks approach.

For the deep web, we think it is necessary to select sites. You have to make a technical and a content analysis one by one, because every one site is different from another. In certain cases some migration may be necessary for long term access, for instance when you have a site with complex architecture, with documents, database and html-pages, and you cannot archive it just like that. So this manual approach is necessary for deep websites.

Regarding robots, we think it is necessary to develop a new kind of robot that would be able to continuously archive sites. For this purpose, it is necessary to focus the crawl on the more documentary part of the site. So we have to compute some more information about the web topography, of the web's information organization. One part of this information, which is very important, is the linking information: who links to whom? For a specific page or a specific site you can calculate the number of incoming links. This is very important, because if there are many links that go to this site, that means that for the web community this site is important in a certain way. It is, to a certain extent, like a bibliographical reference. Search engines like Google use this method to rank results of a query. We could also use information about content evaluation, using other indices and parameters like linguistic information. I will not go into details on this.

With this information at the global scale (the whole web) it will be possible to enhance the selection quality. It will be possible with this information to formulate a crawling and archiving policy for the surface web. That means crawling more fre-



quently to certain important sites; to go there every week or every day if it is necessary. If you can automatically focus the robot on some subset of the web, you don't have to wait six months to come again to these important sites. That allows you to make a more continuous kind of archiving. Of course we will not be able to archive every site and every change of website at the library. That is for sure. But the idea is to have a better follow-up of site-changes. And that means of course that the crawling and the archiving process can be disconnected.

## **:: Archiving the election 2002**

Where are we now at the National Library of France? We have been exploring the way to prepare this new kind of robots for focussing the archive and following the evolution of the sites. We are going to do a snapshot of *.fr* domain in June [2002, ed.] and we have been exploring the deposit track. More than one hundred sites have been contacted and at this moment there are twenty of them that are actually depositing their site in our library. We have analyzed their sites' architecture, the formats they use and we have specified a delivery via FTP or on carrier. The idea is to try to define what kind of procedures and tools we need for the deposit track. This implies of course using the push-model.

In January 2002, we have decided to archive political sites related to the French elections. That is: first we decided to archive sites from candidates and parties. Then we planned to archive sites dedicated to the elections: netcampaign-sites, humouristic sites and so on. We tried to find and archive general sites or other topic-related sites that had part of them, a directory or special pages dedicated to the presidential election. In all, we have found more than five hundred URLs of sites or of parts of sites, or even of documents. Some have been archived twice a week, some each week, some only once, depending on the type of site. In the end we have archived 2.200 sites or parts of sites. We stored all the information about these sites in a database. This means the type of site, the name of the candidate, whether it had a companion discussion list (in this case we tried to take a subscription to this discussion list to archive it) etc. We put this information into the database and then we archived them in an automatic way, using the same tools as Archipol.

- 1 See for instance [http://www.naa.gov.au/recordkeeping/er/web\\_records/intro.html](http://www.naa.gov.au/recordkeeping/er/web_records/intro.html)
- 2 J. Mannerheim, A. Arvidson and K. Persson: 'The Kulturarw3 project - The Royal Swedish Web Archiw3e - An example of "complete" collection of web pages'. Paper presented at the 66th International Federation of Library Associations and Institutions (IFLA) General Conference, Jerusalem, August 2000; see [www.ifla.org/IV/ifla66/papers/154-157e.htm](http://www.ifla.org/IV/ifla66/papers/154-157e.htm).
- 3 More information about the archiving initiatives in different countries at [bibnum.bnf.fr](http://bibnum.bnf.fr).
- 4 See the presentations at the ECDL Web-archiving Workshops; <http://bibnum.bnf.fr/ecdl/>.



Kulturarw<sup>3</sup>

Långsiktigt bevarande av  
elektroniska dokument



## 4 | HET ARCHIPOL-PROJECT

*Henk Druiven, Frank den Hollander, André Keyzer en Gerrit Voerman*

In de loop van 2000 ging het webarchiveringsproject Archipol van start. Hoofddoelstelling was de inrichting van een digitaal archief van de websites van de Nederlandse politieke partijen, en het verkennen van de verschillende aspecten van het archiveren van websites. In dit hoofdstuk worden de resultaten van het archiveringsproject beschreven. Daarbij worden de verschillende fasen in het proces van het archiveren gevolgd:

- het verwerven en downloaden van de sites
- de bewerkingen en aanpassingen
- de opslag, het beheer en de bewaring
- de ontsluiting en toegankelijkheid.

Aan het einde van het hoofdstuk wordt het Archipol-project vergeleken met de Way Back Machine van het Internet Archive.

### **:: Verwerven en downloaden**

In het project zijn de sites van de in de Tweede Kamer vertegenwoordigde partijen gearchiveerd alsmede de sites van de neveninstellingen (met name de jongerenorganisaties), van de kandidaten bij de Tweede-Kamerverkiezingen van mei 2002 en andere sites met betrekking tot deze verkiezingen, ‘tegen-sites’ (bijvoorbeeld een anti-PvdA site) en de sites van partijen die (nog) niet in het parlement zitting hebben. In totaal waren er in de zomer van 2002 in het Archipol-archief bijna 200 verschillende websites opgenomen. Van iedere site zijn meerdere versies gedownload en gearchiveerd.

In de regel wordt een website één keer per maand gedownload. In voorkomende gevallen wordt deze frequentie opgevoerd, bijvoorbeeld in verkiezingstijd, tijdens de kabinetsformatie of na een onverwachte politieke gebeurtenis (zoals de val van het tweede kabinet-Kok in april 2002, en de moord op Pim Fortuyn op 6 mei 2002). In



*De site van de LPF, 8 maart 2002*

principe worden alle pagina's van de site, inclusief afbeeldingen, discussiefora en dergelijke, tot op alle niveaus gedownload. Niet meegenomen worden externe links en de zogeheten *deep web*-informatie van de gearchiveerde websites. Onder deep web-informatie worden de gegevens verstaan die in databases op het web zijn opgeslagen en via een gerichte zoekactie kunnen worden doorzocht; een voorbeeld hiervan is een online catalogus. In de praktijk blijkt het overigens mede door technische tekortkomingen niet altijd mogelijk de complete site volledig geautomatiseerd te downloaden (zie hieronder).

Er is voor gekozen de websites zèlf te downloaden en te verwerken, en niet de webmaster van de betreffende partij te verzoeken om de data aan te leveren en vervolgens alle mutaties door te geven. Door in het verwerkingsproces aan de webmaster een intermediare rol toe te kennen, zou het project te zeer afhankelijk worden van de medewerking van de partijen. Zo kan in geval van grote drukte (bijvoorbeeld tijdens de verkiezingscampagne) er niet voetstoots vanuit worden gegaan dat de webmaster tijd vrijmaakt voor het archiveringsproject.

Uiteraard zijn de politieke partijen wel in kennis gesteld van het project. De reacties waren in het algemeen positief; men zag in dat van bewaring van dit materiaal doorgaans niet veel kwam. Zo is bijvoorbeeld van de eerste versie van de site van GroenLinks en de PvdA uit 1994 bij deze partijen niets bewaard gebleven.

Een tweede reden om de sites zelf te downloaden is gelegen in het feit dat een aantal sites (zoals bijvoorbeeld de CDA-website) wordt opgebouwd via een databasesysteem. Deze sites hebben geen vaste inhoud; de feitelijke opbouw van de pagina's wordt *on the fly* gegene-

**Archipol Archieven:**

- **partijen in het parlement**
- **campagne-archief tweede kamerverkiezingen 2002:**
  - partijen niet in het parlement vertegenwoordigd
  - kamerleden en kandidaat-kamerleden
  - verkiezingen in de media
  - onafhankelijke verkiezingssites
- **gemeenteraads-verkiezingen 2002 in de stad Groningen**

**Archipol Campagne Archief**

**Tweede Kamer verkiezingen 2002: gebruik van internet in de verkiezingscampagnes**

Archipol volgde in de aanloop naar de Tweede Kamerverkiezingen 2002 het gebruik van het internet door de politieke partijen in het Campagne Archief een overzicht van de verkiezingscampagne op internet.

**Toegang en verantwoordelijkheid**

- Oriëntatiepagina's leiden tot het archief, daar u hier een gebruikersnaam en een wachtwoord in voert.
- De gebruiker heeft **verantwoordelijkheid** voor het Archipol archief te gebruiken.

**De verschillende archieven:**

- De partijen die vanaf de verkiezingen 2002 zitting hebben in het parlement (ook inclusief Lijst Pim Fortuyn)
- Politieke partijen die niet in het parlement vertegenwoordigd zijn
- Kamerleden en kandidaat-kamerleden
- Speciale verkiezingssites in de media
- Onafhankelijke verkiezingssites

*Navigatiemogelijkheden op de Archipol-site*

reerd, dus op het moment dat de pagina's door de bezoeker daadwerkelijk worden opgeroepen. Het in zijn geheel aanleveren van dergelijke 'dynamische' sites – inclusief de boomstructuur – is daardoor zeer moeilijk. Bij het archiveren wordt dus in feite een *snapshot* vastgelegd, waarbij de interactieve functionaliteit niet meer beschikbaar is.<sup>1)</sup>

Ten slotte kan niet worden uitgesloten dat webmasters, wanneer hen wordt gevraagd om de site aan te leveren, niet alle pagina's meesturen, of – omgekeerd – ook verouderde of vervallen – niet-gelinkte – pagina's toezenden. Het selecteren van de juiste pagina's (bij de eerste vulling en daarna periodiek) zal voor de beheerders van een site een extra inspanning betekenen. Een garantie dat alleen de 'openbare' pagina's worden aangeleverd, zal door hen dan ook moeilijk te geven zijn. Een controle op volledigheid zou dan weer moeten geschieden aan de hand van een download.

Om al deze redenen is besloten de data 'op te halen', ook al is deze methode arbeidsintensiever. Het voordeel ervan is dat zelf kan worden beslist welke data in welke frequentie worden gedownload.

Voor het feitelijke downloaden van de sites is een eigen programma geschreven. Dit programma (*Archipol.cgi*) is via de website van het project voor vrij gebruik beschikbaar. Verdere ontwikkelingen en uitbreidingen van deze eigen programma-tuur zijn vooralsnog niet voorzien. Tijdens de projectperiode kwam er een andere goed werkende *off line browser* beschikbaar, *HTTrack* geheten.<sup>2)</sup> Voor het downloaden van de websites wordt nu zowel HTTrack als het eigen Archipol-programma gebruikt. De ontwikkeling van HTTrack is nog in volle gang en regelmatig worden nieuwe versies opgeleverd die de technologische ontwikkelingen op het web volgen – zij het altijd wel op enige afstand. Er hoeft zo dus in de toekomst niet geïnvesteerd te worden in het *up to date* houden van de zelf ontwikkelde programmatuur. HTTrack wordt inmiddels ook gebruikt in het archiveringsproject van de Library of Congress, Minerva geheten (Mapping the INternet: the Electronic Resources Virtual Archive).<sup>3)</sup>

Het werken met twee programma's creëert de mogelijkheid om op gezette tijden een controle op volledigheid uit te voeren: is de kopie zoals die wordt opgeslagen in het webarchief een volledige afspiegeling van de actuele site? Daarnaast wordt de gedownloade versie vergeleken met de vorige versie (aan de hand van overzichten van onder meer het aantal bytes en files). Zo is het mogelijk duidelijke afwijkingen (en dus mogelijke lacunes) direct te signaleren. In de praktijk is deze extra controle zeker niet overbodig gebleken. Al in het begin van het project is geconstateerd dat *off line browsers* soms delen van sites niet downloaden.

Het downloaden en archiveren (dat wil zeggen het plaatsen van de gedownloade site in het archief) verloopt vrijwel automatisch. Eerst wordt eenmalig voor elke verschillende website een 'ini-file' aangemaakt. Dit is een parameter-file met een aantal aanwijzingen voor de download, zoals de URL van de betreffende site, het aantal niveaus van de site dat moet worden gedownload, het uitsluiten van externe links en dergelijke. Als eenmaal de ini-file is gedefinieerd, kan op eenvoudige wijze het download- en archiveringsproces worden gestart. Bij het downloaden wordt over het algemeen getracht de server waarop de site is gehost te ontzien door de download zoveel mogelijk in 'daluren' te laten plaatsvinden. Wanneer de site wordt gedownload, maakt het programma zich bekend bij de webmaster van de betreffende site. Indien het downloaden onverhoopt tot capaciteitsproblemen bij een website mocht leiden, kan de webmaster contact opnemen. In de praktijk is dit een enkele keer gebeurd; in overleg met de betreffende webmaster werd vervolgens een tijdstip afgesproken waarop de volgende download kon plaatsvinden.

Het kan voorkomen dat niet alle onderdelen van een site deel uitmaken van de download. Dit kan een aantal oorzaken hebben.

## 1 | Robots.txt

Webmasters kunnen als zij dit willen gebruik maken van het zogeheten 'Robots Exclusion Protocol'. Via dit protocol kan een webmaster delen van de site als zijnde niet-toegankelijk aanmerken voor web crawling robots of voor off line browsers. De delen van een website die niet bezocht mogen worden door een web crawling robot, worden vermeld in een '*robots.txt*'-file die in de *root* van de website is geplaatst. Voorafgaande aan een download of indexering dient de crawlerprogrammatuur eerst deze file te raadplegen.

Deze aanwijzingen worden vanzelfsprekend binnen het Archipol-project opgevolgd: de delen van de site die vermeld staan in een robots.txt-file worden niet gedownload. Aangezien deze onderdelen wel normaal raadpleegbaar zijn voor gebruikers, is hierdoor een deel van de openbare website uitgesloten van archivering.<sup>4)</sup> Wanneer zo'n geval zich dit voordoet, wordt geprobeerd alsnog toestemming van de webmaster te krijgen om deze aanwijzingen ten behoeve van dit project te mogen negeren, teneinde de relevante files te kunnen downloaden en archiveren. Tot dusverre is dit fiat vrijwel altijd verkregen.

## 2 | Interactieve delen van een site

Van interactieve, dynamische aspecten van een website (zoals een antwoord-

formulier, een *chatbox* of de zoekmachine) kan niet meer dan een momentopname worden vastgelegd; die interactiviteit werkt in het archief eenvoudigweg niet meer. Een voorbeeld is de alom bekende Stemwijzer. Deze technische beperking wil overigens geenszins zeggen dat in het archief uitsluitend de tekstbestanden van de websites zijn opgeslagen. Zo wordt bijvoorbeeld ook het ‘Tomaatwerpspel’ van de Socialistische Partij bewaard.

### 3 | Flash en Javascript

Het archiveren van delen van sites waar gebruik wordt gemaakt van nieuwe webtechnologie zoals Javascript of Flash (een toepassing van Macromedia waarin tekst, animatie, afbeeldingen, geluid en interactie worden gecombineerd) verloopt over het algemeen zeer moeizaam.

Met name de sites waarin links naar volgende pagina's zijn opgenomen in een Flash-animatie, zijn zeer lastig te verwerken. Het downloadprogramma (maar ook alle web *search-engines*) kunnen deze links niet herkennen en volgen. Een bijkomend probleem is daarbij nog dat de links in Flash niet makkelijk kunnen worden aangepast. Daardoor is opslag in het archief zeer moeilijk. Indien er geen andere mogelijkheden zijn om deze links te traceren (zoals een *sitemap* – een overzicht van de website-structuur – of html-links op de homepage), dan is het bij de huidige stand van de techniek zeer moeilijk de betreffende onderdelen van de site automatisch te archiveren.

Over dit probleem zijn de webmasters van partijen met een dergelijke site (met name de VVD en de SGP) ingelicht. Het feit dat de site op deze manier ook niet door zoekmachines kan worden ‘gevonden’, kan een overweging voor deze partijen vormen de structuur ervan te wijzigen.

### 4 | Fouten in websites

Een aantal websites bevat inhoudelijke fouten. Deze komen met name in de html-conventies regelmatig voor. Daarnaast kunnen er onjuiste directory-aanduidingen zijn ('//' of '\ ' waar eigenlijk '/' zou moeten staan). De gangbare browsers vallen over het algemeen niet over deze fouten en presenteren de betreffende pagina's wel aan de gebruiker. Download-programmatuur is hier echter wat kritischer, waardoor het kan voorkomen dat eventuele 'foute' links en pagina's niet worden geselecteerd voor downloading. De in het Archipol-project gebruikte download-software kan deze problemen inmiddels wel herkennen en daardoor de files toch downloaden.

## 5 | Oneindige links

Ten slotte vormen zogeheten ‘oneindige links’ een obstakel bij het automatisch archiveren. Dit probleem ontstaat bij programma’s die automatisch steeds nieuwe vervolgpagina’s genereren. Een voorbeeld hiervan is een soort eeuwigdurende kalender, een onderdeel van de site van de Jonge Democraten – de jongerenorganisatie van D66 – waarin vergaderingen en spreekbeurten zijn opgenomen. Bij het downloaden van sites waarop met dergelijke programma’s wordt gewerkt, moet dit deel van de archivering worden uitgezonderd. Dit kan door een aanpassing van de ini-file voor de betreffende site.

### :: Bewerkingen en aanpassingen

30

De gedownloade site wordt in een speciaal hiervoor geschreven programma vergeleken met de eerdere versies van die site die deel uitmaken van het archief. De op dat moment gedownloade files waarvan al een identieke versie in het archief is opgenomen als onderdeel van een site die in het verleden is gedownload, worden verwijderd. Van elke file wordt dus slechts één versie opgeslagen in het archief. Door middel van een *soft link* wordt naar die ‘oude’ file verwezen. Door deze ‘ontdubbeling’ wordt zo efficiënt mogelijk omgegaan met de opslagcapaciteit (zie ook hieronder onder ‘opslag, beheer en bewaring’).

In het ontdubbelingsproces worden tegelijk de interne verwijzingen van een gearcheiverde site (de interne links) aangepast. Deze dienen nu immers naar het webarchief te verwijzen en niet meer naar de oorspronkelijke site. Ook de externe links (dat wil zeggen de verwijzingen naar webpagina’s buiten de site) worden aangepast. Hier wordt een waarschuwing geplaatst dat men het webarchief verlaat en dat deze links naar andere externe pagina’s kunnen verwijzen dan op het moment van archiveren, of dat deze pagina’s zelfs geheel verdwenen kunnen zijn.

Een overzicht van alle noodzakelijke aanpassingen wordt in een metadata-file opgeslagen. Deze file bevat verder onder meer de oorspronkelijke bron (URL), de datum van de download, en informatie omtrent de omvang van de gearcheiverde site. Iedere pagina in het archief wordt voorzien van (extra) metadata-informatie met betrekking tot het archiveringsproces. Daarnaast wordt in het Title-veld van iedere pagina de vermelding ‘ARCHIPOL WEBARCHIEF’ toegevoegd; deze woorden zijn dus boven in het scherm te zien in de blauwe titelbalk van de browser. Via deze vermelding wordt een gebruiker van het archief er nogmaals nadrukkelijk op gewezen dat het hier een kopie in het archief betreft en niet de originele site. Hiermee wordt ook tegemoet gekomen aan de behoefte van de partijen om de gearcheiverde site duidelijk als zodanig herkenbaar te laten zijn; begrijpelijkerwijs wensen zij



te voorkomen dat een oudere versie van hun site voor de actuele wordt aangezien.

Doelstelling van het project is om een zo compleet mogelijke kopie van elke oorspronkelijke site in het archief op te nemen. Naast de – voor de opslag noodzakelijke – aanpassing van de links, de extra metadata-informatie en de vermelding dat het hier een kopie van het origineel betreft, vinden er verder geen bewerkingen meer plaats.

## **:: Opslag, beheer en bewaring**

De gearchiveerde sites worden opgeslagen op de eigen server van Archipol (SUN Ultra 10 onder Solaris 7). Zij worden gesorteerd op website, en daarbinnen weer op datum opgeslagen. Hoewel er bij de opslag – zoals hierboven al is vermeld – wordt ontdebeld, blijft voldoende beschikbare schijfruimte een punt van aanhoudende zorg. De site van alleen al de PvdA telde in september 2001 zo'n 25.000 pagina's met een totale grootte van bijna 650 Mb. De nieuwe versie die in oktober van dat jaar werd gestart, was een stuk kleiner: 900 pagina's en 17 Mb. Het archief is inmiddels zeer omvangrijk geworden: op de 108 Gb datadisks van de Archipol-server stonden half september 2002 ruim 2,5 miljoen files en 150.000 directories met een omvang van 42 Gb.

De beveiliging van de opgeslagen data is een belangrijk aandachtspunt geweest bij de opzet van het archief. Allereerst is er een adequate *back up*-voorziening geïnstalleerd. Verder is er gezorgd voor een goede inhoudelijke beveiliging van de gearchiveerde informatie. Op deze wijze wordt onrechtmatige raadpleging, wijziging, verwijdering of andersoortige manipulatie van de gearchiveerde sites uitgesloten. De authenticiteit en de integriteit van de gearchiveerde data zijn hiermee gewaarborgd.

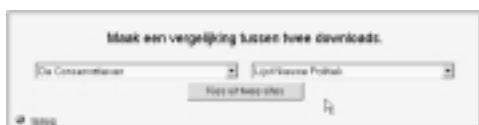
## **:: Ontsluiting en toegankelijkheid**

Aan het downloaden en archiveren van websites kleven niet alleen technische, maar ook juridische problemen. Bij professioneel opgezette websites zoals die van de politieke partijen kan er sprake zijn van een commercieel (grafisch) ontwerp waarop copyright rust. Ook foto's, audio- en videomateriaal en/of artikelen op een site kunnen auteursrechtelijk beschermd zijn. Bij het kopiëren en via een online archief beschikbaar stellen van websites dient men met deze rechten rekening te houden (op de juridische problemen wordt in het volgende hoofdstuk nader ingegaan). Om deze reden werkt het Archipol-project vooralsnog met een gesloten archief. Het bevindt zich weliswaar op het web, maar het is niet openbaar. Het

gedeelte van de Archipol-site met informatie over het project bevindt zich in het openbare domein. Voor het raadplegen van gearchiveerde sites moeten eerst een *username* en *wachtwoord* worden gevraagd. Alleen geautoriseerde gebruikers kunnen toegang krijgen tot het eigenlijke on-line archief. Deze usernames zijn overigens slechts een beperkte tijd geldig. Met de besturen van de politieke partijen vindt overleg plaats over de beschikbaarstelling van het archief en de bijbehorende copyright-kwesties.

De verschillende gearchiveerde websites in het Archipol-archief zijn op diverse wijzen ontsloten. In de eerste plaats kan men ze vinden via de catalogi van de RUG en het DNPP. Verder zijn zij ontsloten in landelijke catalogi als Picarta. De beschrijvingen in de catalogi zijn zodanig opgezet dat het de gebruiker duidelijk is dat het hier om de gearchiveerde versie van de site gaat en niet om de oorspronkelijke, actuele site.

In het archief zelf kan men zoeken op partij (of nevenorganisatie, kandidaat et cetera). Van iedere website kan een overzicht worden opgeroepen van alle gearchiveerde versies. Dit overzicht kan zowel oplopend als aflopend in de tijd worden getoond. Van elke gearchiveerde site wordt de omvang weergegeven in bytes, files, html-files en het totaal aantal gebruikte woorden. Ook kan worden aangegeven welk deel van de site nieuw is ten opzichte van een eerdere in het archief opgenomen versie. Via dit overzicht kan men in één oogopslag een indruk krijgen van de ontwikkeling van een website in de loop der tijd. Ten slotte wordt van iedere gearchiveerde versie van een website een complete sitemap gemaakt en een alfabetische woordenlijst. Via deze sitemap kan men een idee krijgen van de oorspronkelijke opbouw van de website.



Analyse-mogelijkheden

Verder bestaat de mogelijkheid een beperkte inhoudelijke analyse te maken van de versies van de websites die zijn opgenomen. Hiervoor zijn enkele instrumenten ontwikkeld. De gearchiveerde sites zijn alle op woorden doorzoekbaar via de eigen Archipol-zoekmachine, die in het kader van dit project is ontwikkeld. Zo kan onder meer worden nagegaan welke woorden op een site uniek zijn of het meest worden gebruikt, of hoe vaak een woord in de loop van de tijd is

gebruikt in de verschillende gearchiveerde versies van een website (zo komt bijvoorbeeld het woord 'vertrouwensbreuk' geen enkele keer voor in de PvdA-site van 7 mei 2001, maar daarentegen 132 maal in de versie van 6 augustus 2002).

Het navigatiesysteem biedt niet alleen mogelijkheden tot diachroon onderzoek (de ontwikkeling van een bepaalde site door de tijd heen aan de hand van de verschillende versies), maar ook tot synchrone analyse (de vergelijking van sites van verschillende partijen in een bepaalde periode).

## **:: Vergelijking Archipol met de Way Back Machine van het Internet Archive**

Sinds 24 oktober 2001 wordt door het hierboven reeds genoemde *Internet Archive* een archief van websites via het web aangeboden. In dit archief, de *Way Back Machine* geheten, heeft men getracht zoveel mogelijk van het (openbare) Internet op te slaan. In de Way Back Machine zijn overigens alleen websites opgenomen; discussielijsten en dergelijke worden niet gearchiveerd.

De Way Back Machine is gericht op bulkarchivering. Door het volgen van links downloaden *web crawling robots* (ook wel *harvesters* genoemd) elke pagina en daarmee elke site die zij tegenkomen. Deze grootschalige, geheel geautomatiseerde aanpak is primair gericht op kwantiteit: het doel is om een zo groot mogelijk deel van het web te archiveren.

Het nadeel van deze aanpak 'in de breedte' is dat er nog wel eens iets mis gaat, waardoor delen van de gedownloade site ontbreken. Bovendien maakt deze methodiek waarbij een site wordt gearchiveerd op het moment dat het zoekprogramma hem vindt via een link elders, de archiveringsfrequentie van een site tot op zekere hoogte willekeurig. Zo zijn in de Way Back Machine bijvoorbeeld zes versies van de CDA-site opgenomen met als datum 30 mei 1997. In het kader van het Archipol-project worden de websites daarentegen met een vaste regelmaat gearchiveerd. Indien er sprake is van een bijzondere situatie, kan de frequentie op simpele wijze worden opgevoerd.

Met de problematiek van het copyright is bij de opzet van het Internet Archive nauwelijks rekening gehouden. Het gehele archief is openbaar toegankelijk. Inmiddels biedt het Internet Archive eigenaars van sites zelf de mogelijkheid om hun in het archief opgenomen site weer – definitief – te laten verwijderen. Dat betekent dat men er niet zeker van kan zijn dat een gearchiveerde site in de toekomst ook raadpleegbaar zal zijn en blijven.

De in de Way Back Machine opgenomen sites zijn niet ontsloten via een zoekmachine. De enige mogelijkheid om het archief te doorzoeken is via de URL van een site. Een groot nadeel hiervan is dat de gebruiker die geïnteresseerd is in (de ver-

sie van) een website van enkele jaren geleden, de exacte URL van die site op dat tijdstip moet weten. Omdat deze URL niet zelden in de loop der tijd is aangepast en dus niet meer overeenkomt met het actuele adres (als de betreffende site tenminste nog bestaat), zijn oudere versies lastig op te sporen. Binnen het Archipol-project daarentegen zijn gearchiveerde sites ontsloten aan de hand van catalogi en een eigen ontwikkelde zoekmachine en analyse-instrumenten, zoals hierboven reeds werd aangegeven.

34 In tegenstelling tot de brede aanpak van het Internet Archive, legt Archipol zich toe op de bewaring van een specifieke groep websites, namelijk die van de Nederlandse politieke partijen en hun nevenorganisaties. Deze geselecteerde sites worden regelmatig gedownload. In deze kleinschalige benadering is meer oog voor de kwaliteit van de gearchiveerde sites. De werkwijze is dan ook arbeidsintensiever. Zo wordt bijvoorbeeld contact opgenomen met de webmaster van een site waarvan een gedeelte als gevolg van aanwijzingen in de robots.txt-file van archivering is uitgezonderd, en ook wordt er meer aandacht besteed aan de controle van de download. Een duidelijk voordeel van de Way Back Machine is wel dat de programmatuur websites die Javascript en/of Flash bevatten, ook behoorlijk goed weet te downloaden en te archiveren.<sup>5)</sup>

1 Zie ook: F. Boudrez en S. Van den Eynde, Archiveren van websites (versie 1.0), Antwerpen/Leuven, 2002, 29, 34. ([www.antwerpen.be/david.index2.htm](http://www.antwerpen.be/david.index2.htm)).

2 Voor meer informatie zie: [www.httrack.com/](http://www.httrack.com/).

3 Zie de website: [www.loc.gov/minerva/](http://www.loc.gov/minerva/).

4 Uitgesloten kunnen zijn delen van sites, maar bijvoorbeeld ook alle afbeeldingen.

5 Zie ook: Boudrez en Van den Eynde, Archiveren van websites, 32-33.

## 5 | WEBARCHIVEREN EN AUTEURSRECHT

Wilma Mossink

35

Op 22 juli 2002 is het wetsvoorstel 'Uitvoering richtlijn auteursrecht en naburige rechten in de informatiemaatschappij' bij de Tweede Kamer der Staten-Generaal ingediend.<sup>1)</sup> De aanpassing van de Auteurswet moet ervoor zorgen dat ons land qua auteursrechtelijke wetgeving is toegerust voor de informatiemaatschappij en dat de wet voldoet in situaties die in deze veranderde omgeving zijn dan wel gaan ontstaan.

Voordat ik dieper inga op archiveren en de handelingen die daarbij wel en niet zijn toegestaan wil ik eerst iets vertellen over de nieuwe Auteurswet. Dit omdat deze wet het toekomstige handelen bij projecten als Archipol gaat bepalen.

### :: WIPO conferentie 1996

De conferentie van de Word Intellectual Property Organisation (WIPO) – een orgaan van de Verenigde Naties – in 1996 heeft vèrstrekkende gevolgen gehad. Niet alleen inhoudelijk, maar ook territoriaal. In december 1996 zijn in Genève twee verdragen gesloten; het WIPO Auteursrechten Verdrag (Copyright Treaty) en het WIPO Naburige rechten Verdrag (Performances en Phonograms Treaty). Het bijzondere aan deze verdragen is dat het zogenoemde reproductierecht en de beperkingen daarop van toepassing zijn verklaard in de digitale omgeving, zij het slechts in de vorm van een *agreed statement*. Tevens is vastgelegd in deze verdragen dat opslag van een beschermd werk in digitale vorm in een elektronisch medium een verveelvoudiging is.

De Amerikaanse Digital Millennium Copyright Act is de implementatie van het WIPO Auteursrechten Verdrag waardoor de Verenigde Staten dit verdrag hebben geratificeerd. Het Auteursrechten Verdrag is op 6 maart 2002 na toetreding van Ghana als dertigste staat in werking getreden. Het andere verdrag trad in de zomer van 2002 in werking door de ondertekening van het verdrag door Honduras.

De ratificatie door de Europese Unie geschiedt door de lidstaten door de implementatie van de richtlijn 'Harmonisatie van bepaalde aspecten van het auteursrecht en de naburige rechten in de informatiemaatschappij'. In deze richtlijn worden de onderwerpen van de WIPO-verdragen bestreken. De Europese Unie is echter verder willen gaan en heeft ook verschillende onderwerpen willen regelen die niet in de WIPO-verdragen aan de orde komen.

De Auteursrecht-richtlijn, zoals de richtlijn wel genoemd wordt, is het zoveelste product op het gebied van het auteursrecht uit Brussel. Andere voorbeelden zijn de richtlijnen over computerprogramma's, verhuur- en leenrecht en databanken. De richtlijnen komen vaak tot stand na een uitgebreide lobby van belanghebbende partijen, hetgeen de kwaliteit van de regelgeving niet altijd ten goede komt.

De Auteursrecht-richtlijn is op 22 juni 2001 door het Europese Parlement aangenomen. De bepalingen van deze richtlijn moeten binnen achttien maanden zijn omgezet in nationale wetgeving. Omdat de Europese Unie en de lidstaten hebben afgesproken dat zij slechts gezamenlijk toetreden tot de beide WIPO-verdragen, betekent dit dat eerst alle lidstaten de Auteursrecht-richtlijn moeten hebben omgezet in nationale wetgeving. Het is dus begrijpelijk dat de druk vanuit Europa groot is om tijdig met aangepaste wetten te komen.

De Auteursrecht-richtlijn bevat bepalingen over verschillende rechten die de auteursrechthebbers toekomen. Daarnaast regelt de richtlijn de mogelijke wettelijke beperkingen en de juridische bescherming tegen het omzeilen van technische beschermingen die zijn aangebracht.

## **:: Beperkingen en restricties**

Een auteur heeft geen onbegrensde zeggenschap over zijn werk. In verband met de zogenaamde *free flow of information* is nagenoeg iedereen het erover eens dat er enige beperkingen aangebracht zouden moeten kunnen worden op het onbegrensde recht van een maker. In de Berner Conventie staat verwoord dat wetgevers in bijzondere gevallen het verveelvoudigen van beschermde werken kunnen toestaan, mits die verveelvoudiging geen afbreuk doet aan de normale exploitatie van een werk en de wettige belangen van de auteur niet op ongerechtvaardigde wijze geschaad worden. Deze zogeheten 'driestappentoets' komt ook terug in de Auteursrecht-richtlijn.

Het Nederlandse Ministerie van Justitie is voortvarend van start gegaan om de richtlijn te implementeren. Zij heeft daartoe een interactief wetgevingsproces opgezet. Elke burger en belanghebbende kon 76 vragen die betrekking hadden over hoe de richtlijn ingevoerd zou moeten worden, beantwoorden. Op 15 oktober 2001 is het voorontwerp van wet gepubliceerd. Belanghebbenden hadden tot 19 november van dat jaar de gelegenheid hierop te reageren. Na de reacties is het voorontwerp van wet enigszins aangepast en zoals reeds vermeld als wetsontwerp op 22 juli 2002 ingediend. De behandeling van het wetsvoorstel in het parlement zal naar verwachting na de Tweede-Kamerverkiezingen van januari 2003 plaatsvinden.

## **:: Het wetsontwerp**

Het wetsontwerp volgt de drie hoofdonderwerpen van de richtlijn: de rechten, de beperkingen op de rechten, en de technische beschermingsmiddelen.

### **. rechten**

Het auteursrecht geeft de maker zeggenschap over het door hem gemaakte werk. Deze zeggenschap duurt tot zeventig jaar na zijn dood. Artikel 1 van de Auteurswet waarin het auteursrecht wordt gedefinieerd, beschermt de maker tegen ongeoorloofd gebruik door derden van zijn creatie. Het auteursrecht ontstaat op het moment dat een werk vervaardigd wordt. Er zijn geen nadere vormvereisten gesteld. De © *notice* is geen verplichting om auteursrecht te vestigen op het werk.

Het exclusieve auteursrecht bestaat uit twee soorten rechten: de exploitatierechten en de persoonlijkheidsrechten. De exploitatierechten zijn de rechten waarmee toegang verleend wordt tot een werk en waarmee een auteur dan ook geld kan verdienen met betrekking tot zijn werk. De wet noemt twee exploitatierechten: openbaar maken en verveelvoudigen. De Auteurswet geeft geen definities van deze begrippen, enkel nadere omschrijvingen. Openbaar maken is het verspreiden van exemplaren: bijvoorbeeld het uitgeven van een boek, het werk op een website plaatsen of het werk vertonen. Verveelvoudigen is het werk kopiëren, het vertalen of bewerken. Dit geldt niet alleen voor het drukken op papier, maar ook voor opslag in een elektronisch geheugen. De exploitatierechten zijn terug te vinden in de artikelen 12 en 13 van de Auteurswet.

De persoonlijkheidsrechten zijn neergelegd in artikel 25 van de Auteurswet. De rechtvaardiging voor het bestaan van persoonlijkheidsrechten wordt in het algemeen gevonden in de bijzondere band die tussen maker en werk ontstaat, doordat de persoonlijkheid van de maker in zijn werk doorklinkt. Deze persoonlijke, onverbreekelijke band dient beschermd te worden. De persoonlijkheidsrechten geven een auteur het recht zich te verzetten tegen openbaarmaking van zijn werk zonder zijn naam of onder een andere naam dan de zijne. Ook kan hij zich op grond van artikel 25 Auteurswet verzetten tegen wijzigingen in de benaming van zijn werk of andere wijziging in zijn werk, of tegen misvorming, verminking of andere aantasting van zijn werk.

De Auteursrecht-richtlijn schrijft de lidstaten voor dat zij voorzien in een reproductierecht, een recht van mededeling van werken aan het publiek, een recht van beschikbaarstelling en een distributierecht in hun nationale wetgeving. Het verschil tussen het recht van mededeling van werken aan het publiek en het recht van beschikbaarstelling aan het publiek is daarin gelegen dat bij het recht van

mededeling aan het publiek het initiatief bij de bron ligt om het werk ter kennis te brengen aan het publiek, terwijl bij beschikbaarstelling aan het publiek het publiek het initiatief neemt om toegang te krijgen tot het materiaal. Bij het distributierecht is er sprake van verveelvoudiging op een materiële drager. Omdat de Nederlandse Auteurswet geen definities geeft van openbaarmaking en verveelvoudiging, zijn deze begrippen flexibel genoeg om de rechten waarin voorzien moet worden hieronder op te nemen en om ook eventuele nieuwe digitale technieken te omvatten. Zo valt het reproductierecht evenals het distributierecht onder het verveelvoudigingsrecht, en het recht van mededeling aan het publiek onder het openbaarmakingsrecht.

### ■ beperkingen en restricties op de rechten

38 Zoals reeds vermeld heeft een auteur geen onbegrensde zeggenschap over zijn werk. De wetgever heeft er oog voor gehad dat informatievrijheid voor een samenleving een zeer belangrijk goed is. De Auteursrecht-richtlijn geeft lidstaten van de Europese Unie de mogelijkheid om beperkingen en restricties aan te brengen op de exclusieve rechten van de rechthebbenden. Slechts één van deze beperkingen is verplicht, de overigen zijn facultatief.

De verplichte beperking betreft de tijdelijke reproductiehandelingen. Tijdelijke kopieën zonder zelfstandige economische waarde, die van voorbijgaande of incidentele aard zijn, die een integraal en essentieel onderdeel vormen van een technische procédé en die worden toegepast om doorgifte in een netwerk bij rechtmatig gebruik van een werk mogelijk te maken, vallen niet onder het reproductierecht.

Bij het wetsontwerp is Justitie uitgegaan van de bestaande beperkingen. De beperkingen zijn enkel uitgebreid indien er een aangetoonde behoefte aan nieuwe beperkingen bestond. Zo zijn er beperkingen opgenomen ten behoeve van parodieën en ten behoeve van het verwerken van een werk als onderdeel van ondergeschikte betekenis. Ook staan in het wetsontwerp inhoudelijke aanpassingen op bestaande beperkingen.

Nieuw voor voor het publiek toegankelijke bibliotheken, musea en archieven is de beperking dat niet als inbreuk op het auteursrecht wordt beschouwd het door middel van een besloten netwerk beschikbaar stellen van een werk in de voor openbare raadpleging ingerichte ruimten ten behoeve van onderzoek of privéstudie door individuele leden van het publiek. Deze bepaling sluit aan bij de gegroeide praktijk dat het enkel ter beschikking stellen van boeken en tijdschriften voor raadpleging ter plaatse niet als uitlening wordt beschouwd. Deze beperking betreft enkel het raadplegen en heeft geen betrekking op downloaden.

Eveneens nieuw is de beperking voor bibliotheken om ten behoeve van



bewaardoelenden van cultureel en artistiek erfgoed materiaal te mogen verveelvoudigen, de zogeheten archief- of conserveringskopie. De tekst van het wetsvoorstel is geïnspireerd op een voorstel dat door de Federaties van Organisaties in het Bibliotheek-, Informatie- en Documentatiewezen (FOBID) – de koepelorganisatie van alle openbare, universiteits-, speciale en bedrijfsbibliotheken in Nederland – is gedaan tijdens de behandeling van de Databankenwet om te kunnen verveelvoudigen, teneinde zo werken voor verval te behoeden.

Een laatste beperking die hier wordt genoemd is de regeling van de privékopie. Reproduceren voor privégebruik is toegelaten, mits rechthebbenden een redelijke tegemoetkoming ontvangen. De wetgever maakt echter een onderscheid tussen privékopieën waar géén en waarvoor wél een redelijke tegemoetkoming wordt gevraagd. Artikel 16b gaat in op de privé-kopie welke beperkt blijft tot enkele exemplaren en welke uitsluitend dient tot eigen oefening, studie of gebruik van de natuurlijke persoon zonder commercieel oogmerk. Onder artikel 16c valt de privé-reproductie in met name het digitale domein, hoewel het hier niet toe beperkt is. Met betrekking tot de afdracht van vergoedingen is voor alle type dragers waarop verveelvoudigd wordt, aansluiting gezocht bij het bestaande systeem waarin een heffing op de drager ligt. De wetgever geeft aan of een drager binnen het systeem valt.

### **■ technische beschermingsmiddelen**

De bepalingen over de technische beschermingsmiddelen sluiten zo nauw mogelijk aan bij het artikel van de richtlijn. Er is niet gekozen voor een exclusief recht, maar voor een onrechtmatige daadsnorm. De wetgever kan ingrijpen op het moment dat betekenis en ratio van de beperkingen in het gedrang komen. De bepaling maakt het mogelijk dat rechthebbenden verplicht worden dat begunstigten van een beperking de nodige middelen wordt verschaft om van die beperking gebruik te maken.

### **∴ De praktijk**

Welke implicaties hebben deze voorstellen nu voor het archiveren van websites? Laten we eerst eens kijken hoe we een website juridisch kunnen kwalificeren en wat dit betekent voor het archiveren. Een website valt in veel gevallen te kwalificeren als een databank, en dan een databank die geen statisch rustpunt is. Een databank wordt gedefinieerd als een verzameling van werken, gegevens of andere zelfstandige elementen, die systematisch of methodisch geordend, en afzonderlijk met elektronische middelen of anderszins toegankelijk is en waarvan de verkrijging, de controle of de presentatie van de inhoud in kwalitatief of kwantitatief opzicht

getuigt van een substantiële investering. De definitie van een databank komt uit de Databankenwet, de omzetting van de Databankrichtlijn, een richtlijn uit Brussel.

Voordat een website tot stand komt, dient de producent van een databank afspraken te maken met de rechthebbenden van het materiaal dat op de website komt te staan. Flitsende websites bevatten naast teksten vaak ook muziek, en afbeeldingen die al dan niet bewegen. Indien er muziek wordt gebruikt dan is BumaStemra de aangewezen collectieve rechtenorganisatie om hier een licentie voor muziekgebruik te verkrijgen. Voor beeldmateriaal was dit de Stichting Beeldrecht voor beeldende kunstenaars, of de Stichting Burafo voor fotografen. Onlangs hebben deze organisaties aangekondigd samen te werken in een nieuwe stichting genaamd Stichting Visuelen.

Meestal worden de teksten voor websites op maat aangeleverd. Dat betekent dat de schrijvers rechtstreeks benaderd kunnen worden voor het verkrijgen van toestemming voor het gebruik van hun teksten of voor overdracht van hun exploitatierechten, of misschien zelfs wel afstand van hun persoonlijkheidsrechten. Indien foto's op de website geplaatst worden, zullen de geportretteerden in de meeste gevallen wel hun toestemming hebben gegeven. Dit is van belang aangezien wij in Nederland ook het portretrecht kennen.

Wat betreft het aanleggen van *links* heeft de Hoge Raad onlangs beslist dat 'doorlinken' waarbij omvangrijke databestanden gekopieerd worden, niet is toegestaan. In het arrest 'El Cheapo' stelde de Hoge Raad dat databestanden beschermd zijn, ook al maken zij deel uit van een groter geheel. Ook framing is niet toegestaan. Het aanleggen van een hyperlink naar een andere site ter verwijzing is wel toegestaan, omdat dit beschouwd wordt als een soort voetnoot.

Over het gebruik van een databank stelt de Databankenwet dat een rechtmatig gebruiker alle noodzakelijke handelingen mag verrichten ter verkrijging van toegang tot de databank en op het normale gebruik van de inhoud hiervan. De rechten zijn echter niet zo uitgebreid als zij lijken. De rechtmatig gebruiker mag herhaald en systematisch niet-substantiële delen van de inhoud opvragen of hergebruiken. Dit mag echter niet strijdig zijn met de normale exploitatie van de databank, of geen ongerechtvaardigde schade hieraan toebrengen. Een rechtmatig gebruiker van een databank die op enigerlei wijze aan het publiek ter beschikking is gesteld, mag volgens artikel 5 van de Databankenwet zonder toestemming van de producent een substantieel deel van de inhoud van de databank:

- opvragen voor privédoeleinden, mits het een niet elektronische databank betreft;
- opvragen ter illustratie bij onderwijs of voor wetenschappelijk onderzoek, met bronvermelding en voor zover door het niet-commerciële doel

- gerechtvaardigd;  
• opvragen of hergebruiken voor de openbare veiligheid of in het kader van een administratieve of rechterlijke procedure.

De rechten van de rechtmatig gebruiker zijn moeilijk te bepalen. De vaagheid van vele begrippen in de Databankenwet is daar debet aan. Het precieze gebruik zal veelal in een gebruiksovereenkomst worden neergelegd. Onder de rechtmatig gebruiker wordt volgens de Memorie van Toelichting trouwens degene verstaan die een licentie op de databank heeft.

#### ■ archiveren

Wanneer we spreken over archiveren, dan spreken we in feite over meerdere begrippen. We kennen conserveren, preserveren, migreren, converteren en emuleren.

Conserveren is erop gericht om het te bewaren materiaal op de originele drager te behouden. Preserveren is het behouden van de informatie zonder behoud van de originele drager.

Preservering is behoud van de publicatie in ruime zin, ook behoud van de informatie zonder de originele drager.

Migratie is preservering waarbij het digitale materiaal wordt overgezet van de ene hardware/software-configuratie naar de andere, of van een verouderd rakende computertechnologie naar een actuele standaard.

Converteren is omzetten van programma's en/of gegevensbestanden naar de eisen van andere apparatuur of programmatuur. Bij emuleren ten slotte wordt de oude hard- of software met behulp van nieuwe hard- of software nagebootst.

Bij archivering is praktisch altijd sprake van kopiëren. Alleen bij emulatie speelt dit niet. Een kopie is verveelvoudiging en zoals al eerder aangestipt, is het maken van verveelvoudigingen het uitsluitende recht van de auteursrechthebbende. Belangrijk in dezen is dan na te gaan of er een mogelijke beperking bestaat op dit uitsluitende verveelvoudigingsrecht van de rechthebbende. Het hoeft trouwens niet altijd de maker te zijn die de auteursrechthebbende is. De hoofdregel van de Auteurswet stelt dit wel, maar er zijn van rechtswege uitzonderingen hierop. Een maker kan zijn rechten ook hebben overgedragen aan bijvoorbeeld een uitgever of een opdrachtgever, wanneer de maker een werk vervaardigt in opdracht. Een voorbeeld van rechtswege is dat de werkgever rechthebbende is van bepaalde werken in zijn dienst vervaardigd.

Gezien de bepalingen van de Auteurswet en de Databankenwet is het de rechtmatig gebruiker dus niet toegestaan de website in zijn geheel te archiveren. Hier

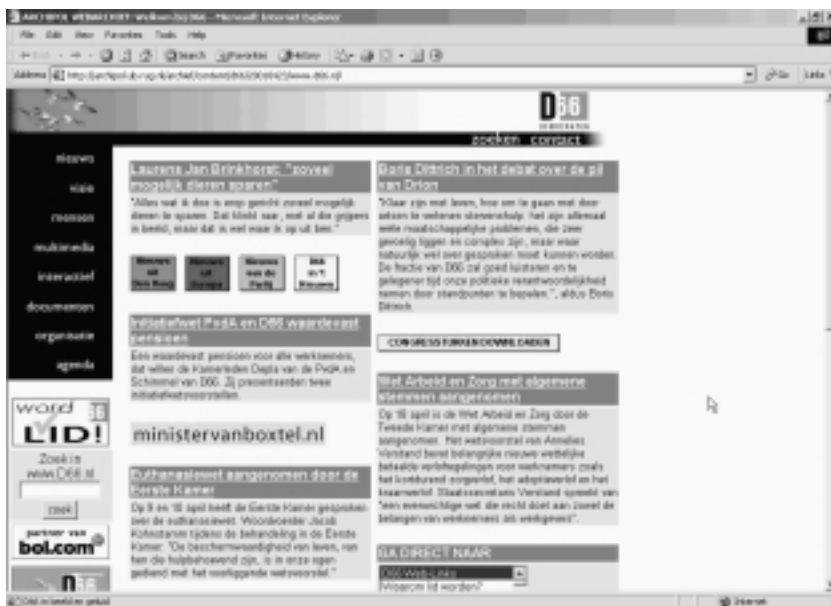
moeten met de producent van de databank dus aparte afspraken over worden gemaakt, bijvoorbeeld door middel van een licentie.

De beperkingen op het auteursrecht staan in de artikelen 15 tot en met 25. Beperkingen op het auteursrecht zijn wat anders dan *'fair use'*. Fair use is een begrip uit de Amerikaanse Copyright Act dat onder toetsing van vier verschillende voorwaarden per casus bepaalt of een gebruiker zonder toestemming te vragen en zonder een vergoeding te betalen aan de rechthebbende, auteursrechtelijk beschermd materiaal kan gebruiken.

## ■ beperkingen voor webarchivering

Er zijn slechts weinig beperkingen in het nieuwe wetsontwerp die van toepassing kunnen zijn op webarchivering ten behoeve van het publiek. Artikel 13a over de tijdelijke kopie is niet van toepassing, omdat archiveren natuurlijk geen reproductie is die van voorbijgaande of incidentele aard is.

Een belangrijk artikel is artikel 16n. De FOBID heeft bij de totstandkoming van de Databankrichtlijn en de uitwerking van deze richtlijn in de Nederlandse wetgeving gepleit voor het opnemen van een bepaling waardoor het bibliotheken toegestaan zou zijn een kopie te maken van auteursrechtelijk beschermd materiaal

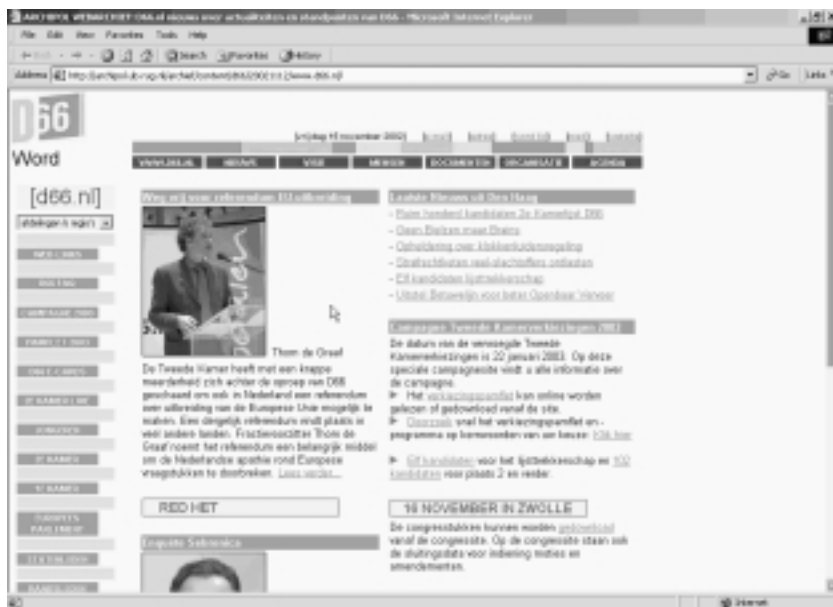


*De site van D66 op 23 april 2002*

omwille van behoud van het materiaal. Een motie van Scheltema verzocht de regering de bepalingen van de Auteurswet, WNR en de Databankenwet weg te nemen die instellingen als bibliotheken, archieven en musea belemmeren bij het preservareren van het culturele, artistieke en publieke erfgoed. De motie beoogde een wijziging in voornoemde wetten, zodat het mogelijk werd beschermd materiaal te verveelvoudigen voor preservatiedoelinden zonder toestemming van de rechthebbende. De motie werd toen kamerbreed gesteund, maar de Minister van Justitie heeft de motie echter niet uitgevoerd: hij gaf er de voorkeur aan om preservatie in Europees verband aan te pakken.

Het verzoek van de FOBID is echter nu wel gehonoreerd, namelijk met artikel 16n. Dit nieuwe artikel stelt dat als inbreuk op het verveelvoudigingsrecht niet wordt beschouwd de verveelvoudiging door archieven of voor het publiek toegankelijke bibliotheken, onderwijsinstellingen of musea die het niet behalen van een economisch of commercieel voordeel nastreven, indien die verveelvoudiging geschiedt met het doel exemplaren van het werk voor verval te behoeden, of als de technologie waarmee het toegankelijk kan worden gemaakt in onbruik raakt, om het werk raadpleegbaar te houden.

Artikel 15h heeft betekenis voor het raadplegen van materiaal. Dit artikel is een beperking op het beschikbaarstellingsrecht. Het door middel van een besloten



De site van D66 op 12 november 2002

netwerk beschikbaar stellen van werken is geen inbreuk indien dit geschiedt door voor het publiek toegankelijke instellingen voor raadpleging van het werk in de voor openbare raadpleging ingerichte ruimten van die instellingen, ten behoeve van onderzoek of privéstudie door individuele leden van het publiek. Wel dient de organisatie die de werken toont, het recht te hebben het materiaal te bewaren en te beheren.

De websites van politieke partijen zijn als een databank te kwalificeren en dus als een werk. Indien artikel 16n in een aangepaste wet terecht komt, zouden het Documentatiecentrum Nederlandse Politieke Partijen en de Universiteitsbibliotheek als onderdeel van de Rijksuniversiteit Groningen als onderwijsinstelling om het werk voor verval te behoeden, een beroep kunnen doen op deze beperking in de Auteurswet. Op basis daarvan zouden zij websites kunnen archiveren en – enkel door middel van een besloten netwerk – ter beschikking kunnen stellen voor raadpleging ten behoeve van onderzoek of privéstudie door individuele leden van het publiek. Als deze beperking niet in de wet terecht komt, zal toestemming moeten worden gevraagd van de producent van de databank om te archiveren.

1 'Aanpassing van de Auteurswet 1912, de Wet op de naburige rechten en de Databankenwet ter uitvoering van de richtlijn 2001/29/EG van het Europees Parlement en de Raad van de Europese Unie van 22 mei 2001 betreffende de harmonisatie van bepaalde aspecten van het auteursrecht en de naburige rechten in de informatiemaatschappij (PhEG L 167)'.

## :: Inleiding

In het Archipol-project is uitgebreid aandacht geschonken aan de wijze waarop de integriteit van de opgeslagen data kan worden gehandhaafd. Er zijn adequate voorzieningen getroffen om het archief jarenlang te kunnen onderhouden en aan te bieden. Het bewaken van de integriteit van de data over een aanzienlijk langere termijn zal een probleem kunnen vormen. Door de snelle ontwikkelingen die plaatsvinden in de informatie- en communicatietechnologie (ICT) kan het platform waarop de data worden aangeboden, op den duur verouderen en wellicht zelfs ontoegankelijk worden, indien er niet tijdig maatregelen worden genomen. Om dit te voorkomen moeten wegen worden gevonden die het mogelijk maken dat digitale databestanden, zoals het Archipol-archief, zonder onaanvaardbaar verlies van informatie ook in een compleet andere ICT-omgeving raadpleegbaar blijven.

Inmiddels worden al vele nationale en internationale initiatieven ontplooid om dit migratieprobleem te onderzoeken. Zo werken in Nederland de ministeries van Onderwijs, Cultuur en Wetenschappen (OC&W) en Binnenlandse Zaken en Koninkrijksrelaties (BZK) samen in het programma 'Digitale Duurzaamheid'.<sup>1)</sup> Binnen dit programma wordt naar praktische oplossingen gezocht in een aantal aan elkaar gerelateerde projecten zoals Testbed Digitale Bewaring, Depot 2000+ en de Taskforce Digitale Duurzaamheid. De ervaringen die binnen het Programma Digitale Duurzaamheid zullen worden opgedaan, zijn beschikbaar voor alle overheidsinstellingen.

In het onderstaande wordt een kort overzicht gegeven van een aantal belangwekkende initiatieven op het gebied van de digitale duurzaamheid.

## 1 | Testbed Digitale Bewaring

In het Testbed-project wordt door de Rijksarchiefdienst en BZK de praktische toepasbaarheid getoetst van de verschillende manieren waarop digitale informatie voor langere tijd kan worden bewaard, zonder dat de toegankelijkheid en de bruikbaarheid in gevaar komen. In een in december 2001 verschenen notitie, *Migratie: context en huidige stand van zaken* geheten,<sup>2)</sup> wordt een overzicht gegeven van de verschillende benaderingen die er op dit moment zijn ten aanzien van digitale bewaring. De tekst van dit hoofdstuk is deels gebaseerd op deze notitie.

In totaal worden er zeven verschillende methodes onderscheiden om digitale

data op langere termijn te bewaren. Als bekendste en meeste gebruikte bewaarstrategie komt *migratie* naar voren. Hieronder wordt door het Testbed het volgende verstaan: 'Het overzetten van bestanden van de ene hardware- of softwareconfiguratie naar de andere.' Uitgaande van deze gehanteerde definitie zal er binnen het Testbed nader onderzoek gedaan worden naar diverse aspecten van migratie, zoals de '*backward*'-compatibiliteit en de interoperabiliteit.

Migratie van data blijkt echter wel arbeidsintensief. Daarnaast is het zeker niet uit te sluiten dat het migratieproces onbekende en ongewenste effecten zou kunnen hebben op de authenticiteit van het bestand. Ook dit aspect zal nader bestudeerd worden in dit project. Uiteindelijk zal dit onderzoek resulteren in een aantal aanbevelingen en adviezen aan de regering en de Rijksarchiefdienst.

Het CAMiLEON-project<sup>3)</sup> (Creative Archiving at Michigan and Leeds: Emulating the Old on the New) van de universiteiten van Michigan en Leeds kiest voor een andere methode. In dit project worden de mogelijkheden van *emulatie* onderzocht als bewaringsstrategie. Via een technologische emulatie wordt op een toekomstig computersysteem een zodanige omgeving gecreëerd dat het oude, oorspronkelijke besturingssysteem inclusief de oorspronkelijke software daarin kunnen opereren. Groot voordeel van deze werkwijze is dat de authenticiteit en de integriteit van de bestanden (in theorie) optimaal kunnen worden gewaarborgd over lange termijn. Op dit moment is er echter bepaald nog geen garantie te geven dat de toekomstige computersystemen in staat zullen zijn om een compleet ander systeem – inclusief alle software – foutloos te laten werken.

### 3 | interPARES

In het interPARES-project<sup>4)</sup> (International Research on Permanent Authentic Records in Electronic Systems) wordt door een groot aantal internationale instellingen gewerkt aan onderzoek naar de theoretische en methodologische aspecten van authenticiteit van digitale data bij bewaring op lange termijn. In tegenstelling tot de eerder genoemde projecten heeft interPARES vooralsnog geen eenduidige bewaringsstrategie, maar worden er algemene conceptionele voorwaarden voor de handhaving van de authenticiteit onderzocht.

De eerste fase (interPARES I) is onlangs afgesloten; de rapportage hierover zal begin 2003 verschijnen. Een vervolgproject (interPARES II genaamd) is inmiddels van start gegaan. Binnen dit project zal naast verder onderzoek naar authenticiteit ook de betrouwbaarheid van digitale data bij langdurige opslag worden bekeken.



Dit project zal lopen tot eind 2006.

## **4 | Archipol**

In het kader van het Archipol-project worden de websites van politieke partijen gedownload en gearcheerd. Deze sites worden opgeslagen op de Unix-server van het project. Naast de archiveringsdata is op deze server ook de ontsluitingssoftware opgeslagen. Bij archiveringsprojecten wordt het materiaal in principe voor de eeuwigheid bewaard. De server zal echter zeker niet voor altijd meegaan. Regelmatig zal vervanging van hardware en/of software noodzakelijk zijn. De installatie van bijvoorbeeld een nieuwe versie van de *operating software* (Unix) zal geen probleem opleveren, maar wat gebeurt er met het Archipol-archief wanneer Unix compleet zou verdwijnen? Dat laatste zal hoogstwaarschijnlijk niet zo snel gebeuren, maar het is wel verstandig om al vroegtijdig rekening te houden met de problematiek van de opslag van de data over langere termijn.

47

### **:: Geen eenduidige richting**

Het bovenstaande overzicht van lopende projecten geeft aan dat er op dit moment nog geen eenduidige richting valt aan te wijzen waarin de bewaring van digitale data in de toekomst vorm zal krijgen. De meest genoemde methoden zijn momenteel migratie en emulatie. Aan beide werkwijzen lijken echter de nodige bezwaren te kleven. Migratie kan ongewenste effecten sorteren, en of emulatie ooit foutloos zal werken is nog maar de vraag. In ieder geval is het duidelijk geworden dat er inmiddels al veel onderzoek plaatsvindt naar oplossingen om digitale data zo ongeschonden als mogelijk voor de toekomst veilig te stellen. Naar mag worden aangenomen, zal dat uiteindelijk resulteren in een technische adequate oplossing voor dit probleem. Binnen het Archipol-project is hier verder niet op ingegaan; afgewacht wordt wat de grote industrieën uiteindelijk aan instrumenten zullen bieden.

Naast de technische problemen die de migratie en/of emulatie van data kan opleveren, zitten er ook juridische haken en ogen aan. Voor deze problematiek wordt hier verwezen naar hoofdstuk 5.

### **:: Tot slot**

In de toekomst zal het zeker noodzakelijk zijn om de gearcheerde data over te brengen naar een ander technisch platform, teneinde de toegang tot het materiaal te waarborgen. Voorlopig is dat nog niet aan de orde. Zolang dit technisch

mogelijk is, zal de Rijksuniversiteit Groningen deze migratie intern uitvoeren.

In het kader van een aantal projecten in binnen- en buitenland wordt er gewerkt aan een oplossing voor het migratieprobleem. Zonder twijfel zullen er hiervoor na verloop van tijd bevredigende methodologische en technische oplossingen beschikbaar komen. Te zijner tijd zal het Archipol-project dankbaar gebruik maken van de bevindingen van deze projecten, en zorgdragen voor de migratie van de data naar een nieuw platform – hetzij door deze migratie zelf uit te voeren, hetzij door aan te sluiten bij al bestaande initiatieven.

1 [www.digiteduurzaamheid.nl](http://www.digiteduurzaamheid.nl)

2 [www.digiteduurzaamheid.nl/bibliotheek/docs/white\\_paper\\_migratie.pdf](http://www.digiteduurzaamheid.nl/bibliotheek/docs/white_paper_migratie.pdf)

3 [www.si.umich.edu/CAMILEON/](http://www.si.umich.edu/CAMILEON/)

4 [www.interpares.org/](http://www.interpares.org/)

In 1992 had vrijwel niemand een idee van wat het World Wide Web was. Tien jaar later, in 2002, lijkt dat onvoorstelbaar. De opkomst van internet heeft de samenleving veranderd. Het kan zelfs degenen die niet tot de *online community* behoren, niet zijn ontgaan. In advertenties en onderaan artikelen in de krant, bij de reclame op televisie of op bestelauto's op de weg: overal verschijnt tegenwoordig het URL, het website-adres van een bedrijf, onderneming, belangengroeperingen of overheidsinstantie. Op deze sites kan worden gekocht, informatie worden gevonden, gereageerd, etcetera. Het publiek dat hiervan gebruik maakt, dijt met grote snelheid uit. In Nederland had in 1995 had nog geen 4% van de meerderjarige bevolking toegang tot internet; in 2001 58%.

Vanzelfsprekend gaan deze ontwikkelingen niet aan de politiek voorbij. Zoals hiervoor al is aangegeven, hebben ook de politieke partijen al jarenlang websites waar bezoekers informatie kunnen vinden en of kunnen reageren of debatteren. Ministeries en gemeenten maken in de informatievoorziening aan het publiek eveneens steeds vaker gebruik van het internet. Andere instellingen zoals de Tweede en Eerste Kamer bieden op hun sites politieke informatie aan. Deze mogelijkheden voorzien kennelijk in een behoefte. Tijdens de campagne voor de Tweede-Kamerverkiezingen van mei 1998 rapporteerde 5,1% van de in het Nederlands Kiezers Onderzoek (NKO) ondervraagde kiezers van het internet gebruik gemaakt te hebben voor informatie over de verkiezingscampagne. Bij de Kamerverkiezingen van 2002 was dit gestegen tot ruim 30%.

Internet maakt nieuwe vormen van politieke participatie mogelijk. Men kan websites van partijen raadplegen, meedoen in discussiegroepen over politiek, debatteren op e-maillijsten van partijen, zoeken naar politieke informatie, of een e-mail versturen aan een politicus. Al deze digitale mogelijkheden hebben geleid tot hooggespannen verwachtingen met betrekking tot politieke participatie en democratie. Optimisten verwachten dat met internet een meer egalitaire, hoogwaardige, directe vorm van democratie binnen handbereik komt. Zo zullen door de traditionele 'oude' media genegeerde politieke groeperingen eenvoudiger hun boodschap voor het voetlicht kunnen brengen. Ook kan de politieke elite in de digitale dialoog gemakkelijker ter verantwoording worden geroepen. In deze optimistische visie zal de toekomst een soort elektronisch Athene te zien geven, met een agora die velen zullen betreden om hun mening kenbaar te maken. Het inherent democratisch karakter van de nieuwe technologie zal groepen mobiliseren die nu

nog niet politiek participeren. Informatie is immers gemakkelijk te krijgen, en deelnemen (reageren per e-mail bijvoorbeeld) kost vrijwel niets. Kortom: in deze gedachtegang leidt internet tot een grotere toegankelijkheid van het politieke domein en tot nieuwe vormen van politiek engagement.

Van deze rooskleurige visie moeten de pessimisten niets hebben. Zij hangen de 'normalisatie-thesen' aan: de bestaande patronen van politieke participatie in het dagelijks leven zullen simpelweg worden gereproduceerd op het internet. In het verleden heeft de introductie van nieuwe informatie- en communicatietechnologie zoals radio en televisie niet geresulteerd in een substantiële toename van politiek engagement, zo luidt hun argumentatie: hoewel burgers zich beter konden informeren, gingen ze politiek niet meer participeren. Van het internet valt dit evenmin te verwachten, al was het maar omdat dit medium vanwege zijn decentrale karakter en *information overload* niet voldoende structuur biedt om gemakkelijk betrouwbare gegevens te vinden. Bovendien zijn bepaalde maatschappelijke groepen van internet uitgesloten, omdat ze de financiële middelen of kennis ontberen om online te gaan. De pessimisten komen op basis hiervan tot de conclusie dat internet de bestaande verschillen in het niveau van politieke participatie zelfs zal versterken.

De pessimisten – of realisten, zo u wilt – lijken het momenteel meer bij het rechte eind te hebben. De introductie van internet heeft vooralsnog niet tot grotere participatie geleid. Uit onderzoek dat het Documentatiecentrum Nederlandse Politieke Partijen samen met het Centrum voor Recht, Bestuur en Informatisering van de Universiteit van Tilburg onder de bezoekers van websites van de Nederlandse politieke partijen uitvoerde, bleek dat maar een relatief bescheiden deel van die bezoekers de behoefte had zich te mengen in een digitaal debat. Het merendeel bezocht de websites van partijen of van instanties die voorlichting over politiek geven om informatie te krijgen. Daarmee lijken de websites – althans voorlopig – niet de digitale fora te worden voor het politieke debat, zoals door sommigen werd gehoopt. Wel krijgen zij een steeds belangrijker functie als schakel in de informatievoorziening tussen partij en achterban. Een iets minder spectaculaire rol misschien dan de profeten van de digitale directe democratie hebben voorzien, maar zeker toch ook belangrijk – zoals weer zal blijken bij de Tweede-Kamerverkiezingen van januari 2003. Het is dan ook van groot belang de websites voor de toekomst te bewaren.

- G. Voerman, A. Keyzer en F. den Hollander, 'Digitale incunabelen. Over het belang van het archiveren van websites, om te beginnen die van de Nederlandse politieke partijen', in: *De Nieuwste Tijd*, no. 15 (dec. 2000), 125-131.
- Gerrit Voerman, André Keyzer en Frank den Hollander, 'Digitale incunabelen. Het belang van het archiveren van websites', in: *Informatie Professional*, 5 (2001), 3 (mrt.), 16-19 ([www.archipol.nl/project/incunabel.html](http://www.archipol.nl/project/incunabel.html)).
- Hans Kuné en Kristien Piersma, 'Het archiveren van websites. Archipol bewaart de digitale geschiedenis', in: *Pictogram*, 3 (2001), 5 (okt./nov.), 4-7 ([www.rug.nl/pictogram/2001-05/interview5-2001.htm](http://www.rug.nl/pictogram/2001-05/interview5-2001.htm)).
- 'Politiek voor de toekomst', in: *Surfcahier*, 6 (2001), no. 30 (nov.), 5 ([www.surf.nl/cahier/Pages/30/30art1.html](http://www.surf.nl/cahier/Pages/30/30art1.html)).
- A.K. Keyzer, F.J. den Hollander en G. Voerman, 'Het Archipol-project. Het archiveren van websites van Nederlandse politieke partijen', in: *Archievenblad*, 106 (2002), 2 (feb.), 32-33.
- Gerrit Voerman, André Keyzer, Frank den Hollander en Henk Druiven, 'Archiving the Web: Political Party Websites in the Netherlands', in: *European Political Science*, 2 (2002), 1 (Autumn); 68-75 (<http://www.essex.ac.uk/ecpr/publications/eps/autumn2002/information.htm>).



| HET SYMPOSIUM  
| IN FOTO'S

*foto's Gerhard Lugard*

►  
Burgemeester van  
Groningen Jacques Wallage  
opent het campagne-archief

Archipol-projectleider  
Gerrit Voerman  
▼



53





◀  
Julien Masanès  
(Bibliothèque Nationale de  
France)

De goedgevulde congreszaal  
van Het Kasteel, met  
vooraan tweede van links  
Wilma Mossink  
▼





**Mischa van Delden** is projectmedewerker elektronische dienstverlening bij de Universiteitsbibliotheek Groningen

**Henk Druiven** is webontwikkelaar bij de Universiteitsbibliotheek Groningen

**Frank den Hollander** is voorlichtingsfunctionaris en webmaster van de Universiteitsbibliotheek Groningen

**André Keyzer** is stafmedewerker elektronische dienstverlening bij de Universiteitsbibliotheek Groningen

**Alex Klugkist** is bibliothecaris van de Rijksuniversiteit Groningen

**Julien Masanès** is leider van het webarchiveringsproject van de Bibliothèque Nationale de France

**Wilma Mossink** is hoofd van de Juridische afdeling van de Open Universiteit Nederland

**Gerrit Voerman** is hoofd van het Documentatiecentrum Nederlandse Politieke Partijen van de Rijksuniversiteit Groningen

In de periode voor de Tweede-Kamerverkiezingen van mei 1998 werd op het Documentatiecentrum Nederlandse Politieke Partijen (DNPP) van de Rijksuniversiteit Groningen onderzoek gedaan naar de manier waarop de politieke partijen in de verkiezingscampagne gebruik maakten van internet. Vrijwel alle partijen hadden in die tijd een website, met informatie over onder meer het verkiezingsprogramma, de kandidaten en de campagne.

Een groot probleem bij het onderzoek vormde het vastleggen van de snel veranderende informatie op de sites. Uiteindelijk bleek er maar één praktische en werkzame oplossing te bestaan om ten behoeve van het onderzoek over de websites te beschikken, namelijk ze allemaal in hun geheel te downloaden. Dit gaf de aanzet tot het Archipol-project: het ARCHIef van websites van POLitieke partijen in Nederland.

Het Archipol-archief wordt inmiddels al enige tijd intensief gebruikt. Kennelijk bestaat er grote belangstelling voor de wijze waarop de partijen digitaal met hun achterban communiceren. Na de dramatische ontwikkelingen in de aanloop naar de Tweede-Kamerverkiezingen van 15 mei 2002 bijvoorbeeld maakten de verschillende partijen volop gebruik van hun websites voor meningen en reacties.

In deze bundel vindt u behalve een uitgebreide beschrijving van het project ook de bijdragen aan het symposium waarmee in april 2002 het archief officieel ten doop werd gehouden.

**archi  
POL**



universiteits  
bibliotheek  
 groningen

Documentatie-  
centrum  
Nederlandse  
Politieke Partijen

ISBN 90-367-1760-4